PHD

Competing Growth Processes with Applications to Networks

Senkevich, Anna

Award date:
2020

Awarding institution:
University of Bath

Link to publication

Competing Growth Processes with Applications to Networks

submitted by

Anna Senkevich

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Mathematical Sciences

January 2020

COPYRIGHT

Declaration of any previous Submission of the Work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

. . . .

Anna Senkevich

Declaration of Authorship

I am the author of this thesis, and the work described herein was carried out by myself personally, with the exception of Section 1.4, Chapters 2 and 4 which contain research that originated from collaboration with my supervisors, Cécile Mailler and Peter Mörters [56].

. . . .

Anna Senkevich

Abstract

In this thesis we define a class of *competing growth processes*, which is a generalisation of reinforced branching processes [26]. The class encompasses different preferential attachment models for networks with fitness such as the Bianconi–Barabási tree (see [15]) and the network of Dereich (see [29]). We analyse the asymptotic behaviour of the largest degree of the network, which corresponds to the largest "family" of our competing growth processes. Apart from networks, our framework also encompasses random permutations with cycle weights (e.g. Chinese restaurant processes), and populations with selection and mutation.

Competing growth processes can be described as a sequence of growing families, which have different birth times and different exponential growth rates. The growth rates are sampled from an i.i.d. sequence of bounded random variables, while the birth times may be random and can depend on the growth process itself. In the most interesting cases the birth times arise from an exponentially growing process so that the largest family at time t arises in competition of the few families born early, which have a longer time to grow, and the many families born late, among which the occurrence of a higher birth rate is more probable.

Our main results show convergence of the scaled size of the largest family at large times to a Fréchet distribution and of the standardised birth time of this family to a Gaussian distribution, in the case where the growth rates are sampled from the maximum domain of attraction of the Gumbel distribution. Furthermore, we compare these results to their counterparts where the growth rates are sampled from the maximum domain of attraction of the Weibull distribution. In this case the scaled size of the largest family also converges to a Fréchet distribution; moreover we obtain the convergence of the fitness of the largest family at large times to a Gamma distribution.

Contents

Chapter 1

Introduction

1.1 Models for real-life networks

Although their impact turned out to be much wider, the main results of this thesis were originally motivated by their application to a model for large networks due to Bianconi and Barabási [15]. This model is one of numerous models for large networks, designed to capture universal properties shared by real-life networks. We introduce the model studied in this PhD after having explained the motivations coming from network science. We begin with a discussion of real-life networks, properties they share (Sections 1.1.1–1.1.3), and associated static and dynamic models (Section 1.1.4). After a brief overview of preferential attachment models, we focus on the Bianconi–Barabási model with fitness (Section 1.2), which was the key motivation for the *reinforced branching processes* (RBPs) framework introduced in [26].

The main contribution of this thesis is the introduction of a new model, called *competing growth processes* (CGPs), which generalises reinforced branching processes (Section 1.3). This new model encompasses different types of preferential attachment networks with fitness, branching processes with selection and mutation, and random permutations with random cycle weights. An overview of examples and a motivational calculation are given in Section 1.3.1 (a detailed discussion of the different examples is reserved for Chapter 4).

In Section 1.3.2 we explain the applications and novelty of our results, namely asymptotic results for the size, fitness and birth time of the largest family at a large finite time, for two classes of bounded-support fitness distributions. The possible directions for further research are outlined in Section 1.3.3, followed by the structure of the thesis in Section 1.3.4. We conclude the introduction with a rigorous definition of our model, statement of our results and assumptions, and a few examples of suitable fitness distributions (Section 1.4).

1.1.1 Why do we need models for large networks?

Developments in computational power and accessibility to high-quality data reignited research interest in the study of complex networks. It became possible to address previously unthinkable questions, and the research interest shifted from properties of small deterministic graphs to statistical properties of large scale random graph models, mimicking properties of real-life networks (see [61] for a survey or [71] for a more extensive overview). Due to size and ever-changing nature, measuring properties of real-life networks directly is an insurmountable task even with the current technologies. Mathematical network models can help to explain the emergence of shared properties in the real-life networks and how these properties affect processes on networks (such as flow of information, spread of diseases, risk propagation, opinion shift and so on, see [71, Chapter 1]). Both of these mathematical topics are fascinating, but we restrict our attention to the former, in particular, we analyse properties of the vertex with the largest degree in a novel class of dynamic networks.

In this section we present a varied selection of large real-life networks, highlighting their similarities, and describe some of the key models suitable for their analysis. This overview is based on the one given by Newman [61], complemented with more recent examples.

1.1.2 Real-life networks

Real-life networks can be split into four loose categories:

1. *Social networks:* groups of interacting people, for example friendship and virtual (online) social networks (e.g. Facebook), collaboration and co-authorship networks;
2. *Information networks:* networks reflecting the structure of information, for example citation networks and the World Wide Web;
3. *Technological networks:* man-made networks designed for resource distribution, for example the Internet, transportation, telecommunication networks and power grids;
4. *Biological networks:* networks observed in biological systems, for example protein interaction networks, metabolic and neural networks (see [61, Section 2]).

In this section we discuss examples and studies of networks from the first three categories. First of all, we look at the evolution of social network studies from Milgram's small-world experiment to Facebook studies, and introduce collaboration networks. From the second category, we discuss citation networks and the Word Wide Web. Finally, we look at the related example from the third category, namely, the Internet. We introduce each network in turn, and then focus on their common properties and the difficulties in analysing them (Section 1.1.3). We refer back to these examples as we discuss different network models in Section 1.1.4. Needless to say that there are many other interesting examples in each of these categories, see [61, 71] for more extensive overviews.

Friendship networks

The importance of social connections cannot be understated: mastering them could shed light on hidden laws and structure governing our society. However, they are notoriously hard to measure: people's interactions constantly change and so do social networks. Questionnaires take time to collect, they are costly and are not always reliable, since everyone interprets social relations differently.

In 1960s, Milgram [57] came up with an ingenious way to probe the distribution of path lengths in an acquaintance network, circumventing the measurement challenge. In a set of experiments, he asked participants to deliver a folder containing documents to a specified recipient, by passing it through somebody that participant knew personally. Many parcels got lost in the course of the experiments, but those that reached the addressee were passed through hands of only six people on average. Interestingly, this experiment inspired a popular notion of the *six degrees of separation*, coined in 1990 by a play-writer Guare [43]. This was one of the earliest attempts to demonstrate the *small-world effect*, which intuitively means that there are relatively short paths between any two pairs of vertices in any network (see [61, Section 2]). We give it a precise definition and discuss it further in Section 1.1.3.

A modern version of Milgram's experiment was conducted by Dodds, Muhamad and Watts [30] in 2001, who used e-mails instead of parcels. Data collected on over 61 thousand individuals from 166 countries, again showed the average number of intermediaries to be six (see [71, Chapter 1]).

Arguably, the appearance of Facebook in 2004 was the game-changer in studies of social interactions, providing ample data on interconnectedness of our society. By 2011 Facebook became the largest online friendship network, with 721 million active users and 69 billion friendship links (see [71, Chapter 1]). In 2012 Facebook reached one billion of monthly active users, and at the time of writing (September 2019), there are 2.3 billion monthly active users according to Wikipedia[1]. From the networks perspective, these active users constitute vertices, connected by an edge if they are friends on Facebook. According to a study by Backstrom *et al.* [7] the average degree of separation on Facebook in 2011 was 3.74, which makes our society seem to have become more interconnected than ever before (see [71, Chapter 1]). However, there

[1]https://en.wikipedia.org/wiki/Facebook

are obvious limitations to how representative of social bonds these online interactions are, not everybody has an account, and some individuals have multiple accounts (some of which are fake). Moreover, there is a limit of 5,000 friendships per account (which already seems to be too large a number for meaningful friendships), which sets an upper bound to the maximum degree and curtails the degree distribution of the social network (see Section 1.1.3).

Collaboration and co-authorship networks

Two additional examples of large social networks with reliable, readily-available data, are movie actor collaboration and mathematics co-authorship networks. In the first example vertices are movie actors and two actors are connected by an edge if they appear in the same film. Similarly in the second example, vertices represent mathematicians, and there is an edge between them if they co-author a paper. The small-world effect is demonstrated by surprisingly small Bacon[2] numbers and Erdős numbers, representing the degree of separation from Kevin Bacon and Paul Erdős in the aforementioned networks respectively (see [61, Section 2]).

Citation network

A citation network is an example of a directed information network, formed of academic articles citing each other. The vertices represent papers and are connected by directed edges, when one article cites another (directed towards the referenced paper). Typically articles only cite existing work, and so all the edges point back in time, making closed loops unlikely. This is why citation networks are considered to be *acyclic*. Availability of reliable data made this network a popular choice of quantitative study from as early as 1926, when Lotka discovered the so-called Law of Scientific Productivity. It states that the number of scientists who wrote k papers is proportional to $k^{-\alpha}$ for some constant α, or in other words, the number of papers written by individual scientists follows a *power-law* distribution (see [61, Section 2]). Such degree distribution seems to be present in many other real-life networks (for example the World Wide Web), and so we come back to this phenomenon in Section 1.1.3.

The Word Wide Web

The World Wide Web is another important example of an information network. It is a directed network of "information-bearing" Web pages, connected by hyperlinks [61, Section 2]. The Web is different from the Internet, which is a technological network consisting of computers linked to each other to transmit information. The Web was invented in 1989 by Tim Berners-Lee. It became available to the general public in the early 1990s, transforming the way information is shared, and has been extensively

[2] http://www.cs.virginia.edu/oracle/

studied since then. Unlike the citation network, the World Wide Web is *cyclic*, that is there exist directed paths that start and end at the same node.

According to Newman [61], studies by Albert, Jeong and Barabási [2, 10], by Kleinberg *et al.* [49], and by Broder *et al.* [21] have been particularly influential. The aforementioned studies conclude that both in-degree and out-degree distributions of the Web appear to follow power laws (see [61, Section 2]). However, the fraction of pages with low in-degree might have been underestimated because the data about the World Wide Web in these studies came from "crawls" of the network (see [53]). Web pages are found by following hyperlinks, which means that the existence of a page is only "discovered" if there is another page with a hyperlink to it (see [21]). This introduces bias into the data, in particular, since a crawl can only cover parts of the Web, pages with more incoming hyperlinks are more likely to be discovered (see [53]).

The Internet

The Internet is a widely studied technological network, consisting of a large number of computers physically connected to each other. The structure of the Internet is usually considered on resolution of either *routers* (special-purpose computers that control the flow of data) or *autonomous systems* (collections of computers within which networking is handled locally, with data flowing over the public Internet, see [61, Section 2]). See for example [22, 39, 24] for some early works on the Internet topology or the Center for Applied Internet Data Analysis[3] (CAIDA) website for extensive, up-to-date measurements on the structure of the Internet.

The Internet is highly decentralised, making it very difficult to determine its structure. Typically, a software, such as *traceroute*, is used to gauge how it is connected. It sends a message and collects the information on the routers visited on the way to its final destination. Putting together these paths enables one to represent the Internet as a graph. However, this is not the original usage traceroute was designed for, and there is a debate about representativeness of and conclusions based on the data collected this way (see [71, Chapter 1]).

1.1.3 Universal properties of real-life networks

Despite describing very different objects, these networks have strikingly many properties in common. The two fundamental properties, shared by our examples (and many other real-life networks) are the *small-world* and the *scale-free* phenomena. The former implies that graph distances between randomly chosen nodes in the real-life networks are relatively small, and the later means that the degrees show a large amount of variability (see [71, Chapter 1]), in other words, a scale-free network does not have a

[3]http://www.caida.org

characteristic scale. We now discuss these two properties in detail, along with a related property of the network's *maximum degree*, the analysis of which motivated our research into the largest family in CGPs. There are other interesting phenomena common to many real-world networks, related to *transitivity*, *mixing patterns*, *community structure*, the size of the *largest component* just to name a few. We direct the interested reader to [61, 71].

Small-world effect

To define the *small-world effect* we consider an undirected network of size n, and define d_{ij} to be the geodesic distance (i.e. the number of edges in the shortest path) from vertex i to vertex j. Let ℓ be the mean geodesic distance between vertex pairs in a network:

$$\ell := \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j}^{n} d_{ij}. \tag{1.1}$$

By convention, if pairs of nodes fall into different components of the network, in other words, if there is no paths connecting them, they are excluded from the average. A network of size n is said to exhibit a *small-world effect* if $\ell \sim \log n$ and an *ultra-small-world effect* if $\ell \sim \log\log n$ (see [61, Section 3]).

The presence of the small-world effect has been verified for various real-life networks (see for example [72]). For some of our examples, we can illustrate it by quoting values of ℓ from a table presented in [61, Section 3]. These numbers, summarised in the table below, suggest that movie actors collaboration network, mathematics co-authorship network, the WWW Altavista directed network and the Internet can all be considered to be small-world networks. Note that it is harder to differentiate between small-world and ultra-small-world networks empirically than analytically.

Real-life network	n	ℓ	$\log n$	$\log\log n$
Movie actors collaboration network	449,913	3.48	13.02	2.57
Mathematics co-authorship network	253,339	7.57	12.44	2.52
WWW Altavista directed network	203,549,046	16.18	19.13	2.95
Internet network	10,697	3.31	9.28	2.23

The table illustrates presence of the small-world phenomena in the real-life networks, where sizes n and mean geodesic distances ℓ are quoted from [61, Section 3], and values of $\log n$ and $\log\log n$ are calculated.

The presence of the small-wold effect in the real-life networks has an important consequence of a relatively fast spread of information, disease, rumour or anything else through the network (see [61, Section 3]). For instance, the number of people that need to contract an infectious disease (for example Ebola) before a local outbreak turns into a global humanitarian crisis is relatively small, due to the interconnectedness of our society, highlighted by the small-world effect.

Scale-free phenomena

A network is said to be *scale-free* when the proportion of nodes with large degree $k >> 1$, denoted by p_k, is proportional to $k^{-\alpha}$ for some constant parameter $\alpha > 0$ as $n \to \infty$. Note that p_k could be interpreted as the probability of a randomly chosen node having degree k. The term “scale-free” is used because the function of degree, $f(k)$, is unchanged (up to a multiplicative factor) under rescaling, that is, $f(ak) = bf(k)$, for some $a, b \in \mathbb{R}$. Alternatively, it is said that degree distribution follows a power law, which is the only functional form with this property (see [61, Section 3]). The degree distribution of p_k of a real-life network can be represented using a histogram plot. These plots are often highly right-skewed, that is, the tail on the right hand side is relatively long, illustrating presence of values which are much greater than the mean. A power-law degree distribution is easily identified by plotting the cumulative degree distribution on a logarithmic scale, which produces a straight line if the distribution is a power law. In this case, the parameter α can be estimated from the slope of the plotted line (see [61, Section 3]).

Scale-free networks are an important and popular topic of study, with the earliest example being Price’s network of citations between scientific papers, published in 1965 [66]. Price empirically established the power law exponent of the in-degree distribution in the real-life citation network to be between 2.5 and 3 (see [66]), and later quoted a more precise value of $\alpha = 3.036$ (see [67]). More recently, power-law degree distributions have been observed in many other real-life networks, including the World Wide Web with in-degree parameter $\alpha = 2.1$ and out-degree parameter $\alpha = 2.7$ (see [61, Section 3]), and the Internet with $\alpha = 2.1$ (see [71, Chapter 1]). Note that there are other common functional forms of the degree distributions identified in the real-life networks, for example exponentials (seen in the power grid and railway networks) and power laws with exponential cutoffs (present in the network of movie actors and some collaboration networks). See for example [61, Section 3] for a more detailed discussion and references.

Scale-free networks exhibit surprising resilience to removal of randomly chosen vertices but they quickly break down under iterative deletion of nodes with the highest degree. *Network resilience* constitutes itself in the robustness of the mean geodesic distance, ℓ, defined in Equation (1.1), and the size of the largest component as a fraction of the network size (used as a proxy measure for *network fragmentation*) to the removal of vertices. These have important consequences for the real-life networks with scale-free property: being great news for targeted vaccination, and bad news for the communication networks under malicious attack.

In their study Albert, Jeong and Barabási [3] simulate behaviour of the Internet under random and targeted attacks. They model the Internet on the level of autonomous systems, assuming a power-law degree distribution with $\alpha = 2.48$ (established by Faloutsos *et al.* [39]). They show that if 2.5% of randomly chosen vertices are removed, corresponding to a high rate of failure of domains, the model remains unaffected. Whereas when 2.5% of nodes with the largest degree are purposefully removed, the average path length, ℓ, triples. Moreover, when 3% of best-connected nodes are removed, the network fragments, implying the complete break down of the connectivity of their model for Internet. This behaviour can be explained by the fact that scale-free networks have a large number of nodes with small degree, insignificant to its overall connectivity, and a small number of very well-connected nodes, whose presence is crucial. Albert *et al.* [3] contrast this with networks with exponential degree distribution, where p_k peaks at the average k and decays exponentially for large values of k (for example the random graph model of Erdös and Rényi [35] and the small-world model of Watts and Strogatz [74], discussed in Section 1.1.4). Such networks are fairly homogenous, with nodes having degrees close to the mean, so a randomly chosen node has a high probability to be well-connected, hence the connectivity decreases monotonically with removal of vertices, regardless of strategy.

Maximum degree

In this thesis we are particularly interested in the *maximum degree* in a large random network. Naturally, this maximum degree is a random variable, and its distribution depends on the size of the network as well as its law [61, Section 3]. It has been shown that in a scale-free network with parameter α, i.e. $p_k \sim k^{-\alpha}$ for large k, we have

$$k_{\max} \sim n^{1/(\alpha-1)}, \tag{1.2}$$

where $k_{\max}$ is the highest degree above which there is less than one vertex on average [61, Section 3]. In particular, Dorogovtsev, Mendes and Samukhin [33] showed that Equation (1.2) holds for Barabási–Albert model [8], a scale-free preferential attachment network model, introduced originally to model the WWW (which we describe in Section 1.1.4). Our results, formulated for the largest family in CGPs (see Section 1.4), show that the maximum degree of the Bianconi–Barabási tree with fitness is in line with Equation (1.2) (see Corollaries 2(i) and 4(i) in Section 1.4).

1.1.4 Historical timeline

It is challenging to understand what processes led the large real-life networks to having properties discussed in the section above. The size and complexity of the networks makes it inappropriate to use deterministic models for their description, hence random network models were introduced. Such random graphs are constructed according to a set of local, probabilistic rules governing the way vertices are connected to each other. According to Van Der Hofstad [71], the field of random graphs was established in the late 1950s and early 1960s, with the paper [36] by Erdős and Rényi from 1960 being thought to have founded the field (see [71, Chapter 1] for more details).

Since then many different models were proposed and analysed. They can be broadly split into two types: *static* (i.e. models of fixed size networks) and *dynamic* (i.e. models of growing networks). We start our discussion of static models with the classic *Poisson random graph* of Erdős and Rényi. We discuss its generalisations: the *generalised (inhomogeneous) random graphs* and the *configuration model*, which constrains a network to have a chosen degree distribution. We also mention the *small-world model* of Watts and Strogatz (see [74]), which gives a good control over mean geodesic distances between vertices. From the dynamic models, we discuss the *vertex copying model* and *preferential attachment models*: *Price's model*, the *Barabási–Albert model* and its extensions, in particular, the *Bianconi–Barabási model with fitness*.

Static random graphs

The Poisson random graph (Erdős–Rényi model)

The *Poisson random graph* is one of the earliest random graph models. It was proposed by Solomonoff and Rapoport in [70], and independently by Erdős and Rényi in [35], after whom it is named. To construct the graph $G_{n,p}$, take n vertices and connect each pair independently with probability p. In a series of papers from 1960s, Erdős and Rényi showed that many properties of the random graph are exactly solvable in the limit of large graph size n, taken holding the mean degree $z := p(n-1)$ constant (see [35, 36, 37]). The graph has a Poisson degree distribution (which is reflected in its name). Newman [61, Section 4] explains that since the presence or absence of edges is independent, the probability of a vertex having degree k, denoted by p_k, is binomial and so, it can be approximated by Poisson in the limit of large n and fixed k,

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \simeq \frac{z^k \mathrm{e}^{-z}}{k!}.$$

Such degree distribution is very different from the highly-skewed power-law distributions of real-life networks (discussed in Section 1.1.1), which highlights the need for other, arguably more elaborate models.

The Erdős–Rényi model captures the small-world effect present in many real-life networks. As Newman [61, Section 4] argues, in a Poisson random graph the mean number of nodes at a distance ℓ from a chosen vertex could be expressed as z^d, for some exponent $d > 0$. To cover the entire network, the value of d needs to be such that $z^\ell \simeq n$. Therefore, an average distance through the network is $\ell = \log n / \log z$, which implies that the presence of the small-world effect. Most other properties of the random graph do not correspond to the real-world networks (see [61, Section 4]) and so, more realistic models were developed since.

The generalised random graph

In the Poisson random graph each vertex plays an identical role, making the graph "egalitarian", unlike the *inhomogeneous* real-world networks, displaying an enormous amount of variability in the roles vertices play (see [71, Chapter 7]). To capture this inhomogeneity Britton, Deijfen and Martin-Löf [20] introduced the *generalised random graph* in 2006. In this model each vertex $i \in \{1, \ldots, n\}$ is equipped with a weight w_i, interpreted as the propensity of the vertex i to have edges. Given the weights, presence or absence of edges is independent. However the occupation probabilities are governed by the weights assigned to the vertices and are therefore different for different edges. There are various implementations of this, with a more general model presented by Bollobás, Janson and Riordan in 2007 [17]. In their version the probability of an edge present between vertices i and j is equal to

$$p_{ij} = \frac{w_i w_j}{\ell_n + w_i w_j},$$

where $\ell_n := \sum_{i \in n} w_i$ is the total weight of all vertices. This leads to vertices with higher weights to having larger degrees, introducing the aforementioned inhomogeneity into the random graph. For suitably chosen weights, this produces graphs with power-law degree distributions (see [71, Chapter 7]).

The configuration model

The *configuration model* presents an alternative approach to producing graphs with non-Poisson degree distributions, representative of real-life networks, by enforcing chosen degree distribution onto the graph. It is constructed as follows. Fix a degree distribution $(p_k)_{k \geq 0}$, where p_k is the proportion of vertices in the network with degree k, and pick an i.i.d. *degree sequence*, $(k_i)_{i \leq n}$, from it. This produces a set of n vertices equipped with some random number k_i of half-edges. Connecting randomly chosen pairs of half-edges generates a graph with specified degree distribution.

A variety of properties of the configuration model have been studied and understood since its introduction in the 1970s (see [12, 58, 63]), with a particular focus on the emergence and size of the *giant component* (i.e. a connected component that contains a positive proportion of the graph's vertices). For instance, an exact condition for the model to possess a giant component has been characterised in terms of the degree distribution $(p_k)_{k\geq 0}$. Moreover, the expected size of that component and the average size of non-giant components both above and below the transition are known. Furthermore, mean numbers of vertices a given distance away from a central vertex and typical vertex-vertex distances in configuration model are also known. The model captures the *high-clustering* property (the extent to which neighbours of vertices are also neighbours of each other [71, Chapter 1]) present in real-life networks with power-law degree distribution (the World Wide Web for instance, see [62]).

The small-world model

The *small-world model*, introduced by Watts and Strogatz [74] incorporates a geographical component into formation of networks, by ensuring that vertices located closer to each other are more likely to be connected. It is constructed from a low-dimensional regular lattice by adding and moving edges to create "shortcuts" (see also [73, 72]). For example, we can start with a *ring*, that is, a one-dimensional lattice of n vertices with periodic boundary conditions [61, Section 6] and connect each vertex to its neighbours that are k or fewer lattice spacings away. We then *rewire* a small fraction of edges, that is, we go through the nk edges and with probability p, we reconnect one end of an edge to a randomly chosen vertex, avoiding self-edges and double edges.

Such procedure maps a regular lattice onto a random graph. For $p = 0$ none of the edges are rewired, and we have a regular lattice with mean geodesic distance ℓ, defined in Equation (1.1), asymptotic to $n/4k$ for large n. For $p = 1$ every edge is moved, producing a random graph with $\ell \sim \log n/\log k$ for $k > 1$ and large n (see [61, Section 6]). Note that, the former graph does not exhibit the small-world effect, whereas the later does. The degree distribution of the small-world model is given by the following expression

$$p_j = \sum_{n=0}^{\min(j-k,k)} \binom{k}{n} (1-p)^n p^{k-n} \frac{(pk)^{j-k-n}}{(j-k-n)!} \mathrm{e}^{-pk}, \quad ,\text{for } j \geq k$$

and $p_j = 0$ for $j < k$ (see [11]), implying that there are no degrees smaller than k which is not representative of the real-life networks (see Section 1.1.3).

Dynamic random graphs

The static models described above are *phenomenological models*, i.e. they are used to recreate specific properties, such as degree distribution or small-world phenomenon for example. In contrast, dynamic models, constructed by gradual addition of vertices and edges, attempt to explain the emergence of such properties, by focusing on rules governing growth of the real-life networks. Here we discuss models aimed at explaining the origin of highly skewed degree distributions. Before focusing on models with preferential attachment mechanism, we briefly outline the vertex copying model, which is an alternative dynamic model for capturing the power-law degree distribution.

Broadly speaking, *preferential attachment* refers to "the rich get richer" phenomenon, which from the networks perspective translates into nodes with already relatively high degrees being more likely to attract new edges, thus increasing their degree further. We begin our discussion of preferential attachment models with a summary of different preferential attachment rules with and without fitness, outlined by Bhamidi [14]. For simplicity, we illustrate them with tree models, rather than for general networks. We then give a historical overview of the development of preferential attachment networks as a modelling tool for real-life networks. Starting from the cumulative advantage model of Price [67], introduced for modelling citation networks, we proceed to the highly influential model of Barabási and Albert [8], inspired by the World Wide Web. Finally we talk about extensions of the Barabási–Albert model, including the Bianconi–Barabási model with fitness, which is of a particular interest to us and is discussed in greater detail in Sections 1.2 and 4.2.1.

The vertex copying model

The *vertex copying model* was proposed by Kleinberg *et al.* [49] as a model for the World Wide Web. The graph is constructed by addition of vertices to the network, and equipping them with directed edges which are either randomly chosen or copied from a randomly chosen existing vertex. In the basic version, a new vertex copies some or all of the edges from a randomly chosen node in the network, hence connecting to some or all of the vertices that the selected node is connected to (see [61, Section 7]).

Kleinberg *et al.* [49] argue that the copying mechanism is representative of "content-creation" on the Web and explains its statistical properties, such as the power-law degree distribution, for example. In particular, the probability of an edge from a randomly chosen vertex leading to a specific vertex with in-degree k is proportional to k, implying that the in-degree of a vertex increases with rate proportional its current value. Kumar *et al.* [52] established that the vertex copying model has a power-law degree distribution with the exponent $\alpha = {}^{(2-a)}/_{(1-a)}$, where a is the ratio of the number of edges with randomly chosen targets to the ones with copied targets. For a between 0 and $1/2$, corresponding to the case where most targets are copied, the exponent α lies between 2 and 3, which is in line with the values observed in many real-life networks (see [61, Section 7]).

Models of preferential attachment

Summary of preferential attachment rules for trees

Bhamidi [14] describes a general *preferential attachment mechanism* for trees as follows. At time step $t = 1$, there is one node, labelled 1. For $n \geq 1$, at time step $t = n + 1$, the $(n + 1)$st node is added to the tree and connected with one edge to one of the existing n nodes. This node is chosen with probability proportional to some specified *attractiveness* function $f(v, n)$ of node v at time n. Many different variations of this attractiveness function have been studied, and we list a few of them below on an example of a directed tree (see [14] for a more extensive summary).

The directed preferential attachment tree can be thought of as growing downwards, i.e. edges are directed away from the root, so the first node has in-degree zero, and all other nodes have in-degrees equal to one. The overall degree of any vertex apart from the root is therefore $D(v, n)+1$, where $D(v, n)$ is the out-degree of the vertex v at time n. Fix $a \geq 0$ and an integer $A > 1$. For preferential attachment models with fitness we also need to fix a probability measure μ on $\mathbb{R}^+$ and assign to each vertex v its fitness F_v chosen independently from μ. Below is a non-exhaustive list of attractiveness functions $f(v, n)$ for different preferential attachment models, summarised by Bhamidi [14]:

- *Linear Preferential Attachment model*: $f(v, n) = D(v, n) + 1 + a$, for fixed $a \geq 0$;
- *Sub-linear preferential attachment*: $f(v, n) = (D(v, n) + 1)^\alpha$, for $0 < \alpha < 1$;
- *Preferential attachment with a cutoff*: $f(v, n) = D(v, n) + 1$ if $D(v, n) \leq A$ and $f(v, n) = A + 1$, if $D(v, n) > A$;
- *Preferential attachment with additive fitness*: $f(v, n) = D(v, n) + 1 + F_v$;
- *Preferential attachment with multiplicative fitness*: $f(v, n) = (D(v, n) + 1)F_v$.

In our research we focus on the preferential attachment with multiplicative fitness, to keep it in line with the Bianconi–Barabási model, which we discuss below. These models can be easily extended to more general networks by allowing an arbitrary number of connections for incoming nodes. An example of this is the preferential attachment network of Dereich [25], which we introduce in Section 4.2.2 and study in detail in Section 4.4.2. Below we present a short historical overview of the preferential attachment models.

The cumulative advantage model

According to Newman [61, Section 7], in 1965 the "physicist-turned-historian-of-science" Price [66] described what is thought to be the first scale-free network. He studied the network of citations between scientific papers and found that both in-degrees and out-degrees, corresponding to the number of times a paper has been cited and a number of other papers a paper cites, have power-law distributions (see [66]). Price's work built upon ideas developed in the 1950s by Simon [69], who showed that power laws emerge when "the rich get richer", i.e. when the amount one gets increases with the amount one already has. Price called this phenomenon *cumulative advantage*, but it is often referred to as *preferential attachment*, the phrase coined by Barabási and Albert [8]. The cumulative advantage in scientific paper citations is easy to justify in a qualitative way. Papers, which are more prominent already, are likely to attract more attention in the future.

In Price's model new vertices are gradually added to a directed network, representing a citation network. Each vertex represents a scientific paper, and is equipped with some fixed out-degree, corresponding to the number of papers it cites. The out-degrees are allowed to vary between vertices, but the mean out-degree is kept constant over time. The *cumulative advantage* manifests itself in the new edges being connected to randomly chosen existing vertices, with probability proportional to their in-degrees, denoted by k. This represents a newly published paper citing an existing one. Since each vertex has initial in-degree zero, a constant offset term k_0 is introduced. Price set k_0 to one, interpreting the publication of a paper as its first citation. Hence, the probability that a new edge attaches to any of the vertices with in-degree k is

$$\frac{(k+1)p_k}{\sum_k (k+1)p_k} = \frac{(k+1)p_k}{m+1} ,$$

where p_k is the fraction of vertices in the network with in-degree k, and m is the mean in-degree of the network (so that $\sum_k p_k = 1$ and $\sum_k kp_k = m$ by definition).

In the large-n limit, the in-degree distribution has a power-law tail with exponent $\alpha = 2 + 1/m$, in other words, $p_k \sim k^{-(2+1/m)}$ as $n \to \infty$ (see [61, Section 7]). The exponent α does not depend on k_0, implying that the choice of the offset is not important. For $m \geq 1$, which corresponds to the case when a paper cites more than one existing works on average, the exponents lie in the interval between 2 and 3, which is in line with the values observed in real-world networks (as discussed in Section 1.1.1).

The Barabási–Albert model

In 1999 Barabási and Albert [8] proposed a dynamic network model for the World

Wide Web, built on a concept similar to Price's cumulative advantage, which they refer to as "preferential attachment". The network grows by addition of vertices which connect to $m \geq 1$ existing ones, chosen randomly with probability proportional to their degree:

$$\frac{kp_k}{\sum_k kp_k}\,.$$

The assumption of linear preferential attachment was justified by further studies of citation networks, the Internet, and actor and scientist collaboration networks (see for example [47, 60]). Note that the initial number of edges m is fixed for all incoming vertices, hence unlike Price's model the initial degree is not allowed to vary around the mean initial degree m. The new vertex automatically has a non-zero initial degree, translating into a non-zero probability of attracting links from vertices introduced into the network at a later stage. Hence removing the directionality of the edges helps to avoid the question of how a paper gets its first citation or a website gets its first link, which Price had to resolve using an offset term. However, it makes the Barabási–Albert model less realistic than Price's model, since both citation networks and the Web are directed graphs.

The Barabási–Albert model has been solved exactly in the limit of large graph size, see for example Krapivsky, Redner, and Leyvraz [51] or Dorogovtsev, Mendes and Samukhin [33]. Most importantly, the Barabási–Albert model leads to power-law degree distribution with a single exponent 3, i.e. $p_k \sim k^{-3}$ in the limit of large k (see [18]). Another important result, shown by Krapivsky and Render [50], is the correlation between the age of vertices and their degrees. The vertices added to the network earlier have substantially higher expected degrees than vertices added at a later stage, with the emergence of power-law degree distribution being due to the earliest vertices (see [61]). Adamic and Huberman [1] show that in the Web data age and degrees are not correlated, and argue that this highlights the unsuitability of the Barabási–Albert model for representing the Web, whose dynamics is more complicated than can be captured in this model. An extension of the Barabási–Albert model, in which age and degree are not correlated, has been proposed by Bianconi and Barabási [15]; we discuss it below and in Sections 1.2, 4.2.1 and 4.4.1.

The Bianconi–Barabási model and other preferential attachment models with fitness

To make the Barabási–Albert model more realistic, various extensions were introduced (see [9] for an extensive overview). Many of them can be categorised according to the type of the preferential attachment rule, similar to the ones we listed for trees. For example, the model with linear preferential attachment was studied by Dorogovtsev, Mendes and Samukhin [33] and Krapivsky and Redner [50]. Krapivsky, Redner and Leyvraz [51] examine the sub-linear preferential attachment. Allowing the mean degree

m to change over time was considered by Dorogovtsev and Mendes [32, 31], who tried to incorporate the increase in the average degree of a vertex in the World Wide Web.

A qualitatively different class of preferential attachment models explores the idea of varying intrinsic worth of vertices, referred to as *fitness value*, as a governing component in the growth of the network. The inspiration for these models come from the work of Adamic and Huberman [1] who have shown that unlike the model of Barabási and Albert, the World Wide Web does not have the correlations between age and degree of vertices. They argue that this is due to the degree of vertices being a function of their intrinsic worth: some websites are more "useful" than others and so they gain links at a faster rate.

To incorporate this idea, Bianconi and Barabási [15] extended the Barabási–Albert model, by equipping each vertex $i = 1, \ldots, n$ with a random *fitness*, F_i, representing its ability to attract new links (see [15, 16]). These fitness values are chosen from a specified distribution μ and once assigned to a vertex they remain unchanged. In the Bianconi–Barabási model the probability of attachment is proportional to the product of the fitness and degree of the vertex, $F_i k_i$, for $i = 1, ..., n-1$, making it a preferential attachment model with multiplicative fitness. Including fitness leads to a much richer behaviour, giving the fit late-comers a chance in competing for links against older vertices. In Section 1.2 we discuss the role of the fitness distribution μ in governing growth dynamics of the network.

There are several variations on the fitness theme. For instance, a model with fitness but without preferential attachment has been studied by Caldarelli *et al.* [23]. It has been shown to produce power-law distributions under specific assumptions. A directed version of the Bianconi–Barabási model where the fitness F_n contributes additively to the attachment probability have been studied by Ergün and Rodgers [38]. They found that for a suitable choice of parameter values, the power-law degree distribution is preserved, and the value of the exponent may be affected by the choice of distribution μ from which fitness values are drawn (see [61, Section 7] for a few more examples).

1.2 The Bianconi–Barabási model with fitness

We now come back to the aforementioned Bianconi–Barabási model with fitness and discuss its behaviour in detail.

1.2.1 Definition and mapping to Bose gas

The model was proposed in 2001 by Bianconi and Barabási [15], in an attempt to capture the growth dynamics of the World Wide Web, economic networks and citation networks through their competition for links, be it hyperlinks, connections to customers or citations. Empirical studies on these networks suggest that the addition of websites, the emergence of new companies or the publication of new papers happen at different rates for different nodes. These nodes "self-organise into a complex network" whose structure evolves as a result of this competition (see [15]). The varying ability of the nodes to attract links, due to usefulness of the websites' content, quality of products and the novelty of research, is encapsulated by a single fitness parameter assigned to each node.

Using our notation we can describe the Bianconi–Barabási model with fitness as follows. At each time step one new node, n, is added to the network. This node n is equipped with a random fitness value F_n, sampled from a specified distribution μ, independently of everything else. Once a fitness value is assigned to a node it stays fixed. The incoming node connects to m existing vertices, such that it chooses to connect to a node i with probability proportional to the product of fitness and degree of the node i, given by

$$\frac{F_i k_i}{\sum_{j \le n} F_j k_j},$$

where k_i is the degree of the ith node. Thus the incoming nodes prefer to connect to nodes which are not only more "popular" but also "better", which could be interpreted from the perspective of the World Wide Web as new-comers favouring more visible websites and websites with more useful content (see [15]).

Incorporating fitness values into the preferential attachment network model leads to much richer behaviour. Bianconi and Barabási [15] conjectured a possibility of three different phases determined by the choice of the fitness distribution μ. Borgs, Chayes, Daskalakis and Roch [19] present rigorous analysis of these phases, which they describe as:

- the *first-mover-advantage* phase, which arises for flat fitness distributions, leading to the power-law behaviour similar to the linear preferential attachment model without fitness;
- the *fit-get-richer* phase, in which vertices with higher fitness accumulate new edges with a significantly faster rate than the less fit vertices;
- the *innovation-pays-off* phase, in which a constant fraction of the links continuously shifts to larger fitness values, "escaping to infinity".

Borgs *et al.* [19] characterise the growth dynamics of each phase and the properties of fitness distribution μ, needed for each of these three phases to occur. Moreover, they prove the existence of the *innovation-pays-off* phase, in which a proportion of the mass in the degree-weighted fitness distribution condenses in a maximal fitness. This phase was predicted by Bianconi and Barabási [15], who argue that in the thermodynamic limit ($t \to \infty$), their fitness model maps into a Bose gas, where energy levels correspond to a function of fitnesses, and noninteracting particles on different energy levels correspond to nodes connected by an edge.

By comparing the physical and networks models Bianconi and Barabási conjecture a possibility of Bose–Einstein *condensation*, which they define as "the fittest node acquiring a finite fraction of the links, independent of the size of the network" (see [15]). However this definition seems to be too strong and presence of the so-called *macroscopic condensation* for bounded regularly varying functions was disproven in [26] by Dereich, Mailler and Mörters, who try to answer questions like "When did the nodes that form the condensate enter the system?" and "How many nodes contribute to condensate?". These questions motivate the analysis of asymptotic properties of the node with the maximal degree, which is the main interest of this thesis.

1.2.2 Techniques for the analysis of preferential attachment with fitness

Dynamics of random graphs can usually be analysed using a *mean-field method* (see [9] for the analysis of the Barabási–Albert preferential attachment model), where the average growth of the system is used to predict growth dynamics of individual vertices. However, due to additional randomness introduced through fitness values, the growth dynamics in the preferential attachment networks with fitness is too subtle for this technique to give meaningful insights. Coupling with generalised Pólya urns and embedding into continuous time branching processes are two very important techniques that are capable of capturing the more intricate growth dynamics. They form the basis of our analysis, and we introduce them below.

Coupling with generalised Pólya urns

An approach suitable for the analysis of preferential attachment with fitness is introduced by Borgs *et al.* [19], who use coupling with *generalised Pólya urns* (see [46]). Borgs *et al.* argue that although initially derived for the analysis of an undirected tree, the technique could be extended to directed scale-free graphs (see [19]).

The classical Pólya's urn model is a random iterative process that describes the contents of an urn containing balls of two different colours. At each time step a ball is drawn from the urn. It is then returned back and one more ball of the same colour is added to the urn. In the preferential attachment tree, the degree of the first two vertices evolve like a Pólya urn. In the generalised version of Pólya's urn model, an "activity parameter" is assigned to each colour, and it determines the likelihood of a ball being drawn (which is analogous to the fitness parameter in preferential attachment networks with fitness). The number of colours in this scheme is allowed to be arbitrary, although it remains finite. At each time step, a randomly chosen ball is returned alongside a random number of balls of each colour, such that the distribution of this "random update vector" is determined by the colour of the picked ball [19].

To study asymptotic properties of the preferential attachment networks with fitness, Borgs *et al.* [19] couple the growth process with specially-designed generalised Pólya urn models where colours are associated with the cumulative degree of all the vertices of a particular fitness. Such a representation is relatively straightforward in the case where the fitness distribution is concentrated on a finite number of atoms, and so the non-trivial generalisations of classic results can be read off from the coupling. This is not the case for more general fitness distributions, including continuous distributions, where an infinite number of colours in the Pólya urn model would be needed for this approach. Since the behaviour of the generalised Pólya urns in the infinite case, is not well-understood, instead, various truncation techniques are used to map the dynamics of the network to a finite urn process (see [19] for more details).

Embeddings into continuous time branching processes

Athreya and Karlin [5] developed techniques for embeddings of urn schemes into continuous time branching processes, which were later advocated by Janson [46]. Bhamidi [14] takes this idea further and directly embeds discrete time networks into continuous time branching processes. He can then apply some of the classical results (see [45]) to complete the analysis and gain insights into the limit behaviour of the network. This approach gives asymptotics for various statistics associated with different models of preferential attachment networks. The technique is particularly effective when applied to trees, giving results on its characteristics such as height, asymptotic degree distribution, degree of the root and maximum degree (see [14]).

The embedding relies on *multi-type branching processes* (see [45]), with the main idea being that each node "reproduces" (i.e. connects to an incoming node) independently at ages according to a random point process ξ. The key assumption for the convergence theory of Crump–Mode–Jagers processes (see for example [59]) is the existence of a *Malthusian rate of growth parameter*, i.e. there exists $\lambda > 0$ such that

$$1 = \int_0^\infty \mathrm{e}^{-\lambda s}\mathbb{E}\xi(\mathrm{d}s). \tag{1.3}$$

One of the conclusions of Bhamidi's analysis is that asymptotic properties of the trees are governed predominantly by the Malthusian parameter (see [14]). In particular, when the Malthusian parameter exists one can use the results from Nerman [59] to show that the asymptotic degree distribution of the Bianconi–Barabási tree with fitness converges almost surely to

$$p_k = \int_0^\infty \frac{\lambda}{kf+\lambda} \prod_{i=1}^{k-1} \frac{if}{if+\lambda} \mu(\mathrm{d}f), \quad \forall k \geq 1. \tag{1.4}$$

For bounded fitness distributions with the essential supremum set to one, it is easy to show that λ lies between 1 and 2, and from Equation (1.4) it follows that $p_k = k^{-(1+\lambda)+o_k(1)}$ (see [55, Chapter 3]). This implies that the Bianconi–Barabási tree with bounded fitness is scale-free with parameter $1+\lambda \in (2,3)$. Equation (1.2) suggests that the maximum degree of the Bianconi–Barabási tree with bounded fitness should be of order $n^{1/\lambda}$. We confirm this with our Corollaries 2(i) and 4(i) in Section 1.4 which we prove in Sections 2.2.5 and 3.3.2.

1.2.3 Reinforced branching processes (RBPs)

Building upon the aforementioned techniques, Dereich, Mailler and Mörters [26] study a class of branching processes in which individuals are equipped with a fitness value. These processes are referred to as *reinforced* branching processes, because particles of the same type reproduce and so their presence is "reinforced" (see [65]). The advantage of looking at RBPs is that they encompass a number of different objects, including Bianconi–Barabási trees with fitness, Pólya urns and also population growth with selection and mutation. The downsides of the RBPs are that the number of offspring in a single birth event is at most two and from the networks perspective RBPs do not easily extend beyond trees.

Construction

The easiest way to describe the RBPs is as a growing population with selection and mutation. Individuals are immortal and equipped with fitness value which determines their rate of reproduction. The fitness value once assigned to an individual remains unchanged. In other words, mutation only happens at birth (this is analogous to the inertness property of the energy levels of Bose gas described in Section 1.2.1, see [15]). Individuals are organised into *families* according to their fitness value, i.e. a family is a collection of all individuals in the population with the same fitness value (see [26]).

Let μ denote a suitable fitness distribution and β, γ denote the probabilities of mutation and selection respectively, such that $1 \leq \beta+\gamma \leq 2$ (which allows to have one or two offspring at each birth event). The process starts with one family consisting of one individual, equipped with a random fitness drawn from the distribution μ. The individual reproduces at a rate proportional to their fitness. At each birth event a *selectant* is born with probability γ (i.e. the offspring inherits the fitness value of the parent), and a *mutant* is born with probability β (i.e. the offspring gets a new fitness value drawn from μ, and so a new family is established). Note that two individuals are born in a single birth event with probability $(\gamma + \beta) - 1$.

Branching processes with selection and mutation

If $\gamma + \beta = 1$, there is only one child born at a birth event: either a selectant or a mutant. This is similar to Kingman's model (see [48, 28]). This gives a stochastic *house-of-cards model*, where the name comes from the fact that mutation causes the complete loss of genetic information, and so "the genetic house of cards collapses" (see [26]). The parallel with branching processes allows one to use the wealth of results and techniques for the analysis of the model (see for example [6]). It can be shown that the growth of a family n is a *Yule process* with intensity γF_n.

Recall that the *Yule process* $(Y_t : t \geq 0)$ with rate η is a process of immortal particles, which starts with one particle and at any time every particle independently gives birth to a new particle with rate η. At any time $t \geq 0$, Y_t is the number of particles alive at time t. This process is characterised as follows: let τ be an exponential random variable of parameter η, then $Y_t = 1$ for all $t < \tau$, and for all $t \geq \tau$, $Y_t = Y^{(1)}_{t-\tau} + Y^{(2)}_{t-\tau}$ where $Y^{(1)}$ and $Y^{(2)}$ are two independent copies of Y (see for example [6, 55] for a discussion and interesting results about the process).

The appearance of new mutants in the system is a Crump–Mode–Jagers process (see for example [59]), and a mutant produces new mutants as a Cox process (i.e. a Poisson process with a random intensity measure). In this setting the Malthusian parameter, defined in Equation (1.3), plays an important role in determining the growth dynamics of the population. Combining Equation (1.3) and the model definition, it can be shown that the Malthusian parameter exists if and only if $\frac{\beta}{1-\beta} \int \frac{f}{1-f} \mu(\mathrm{d}f) \geq 1$ (see [26]). If this condition fails, the classical convergence theory of Crump–Mode–Jagers processes fails, which leads to emergence of condensation (see [26]).

The Bianconi–Barabási tree with fitness

Alternatively, if we set $\beta = \gamma = 1$, at each birth event there are always two offspring: a selectant and a mutant. This recovers the Bianconi–Barabási tree with fitness, embedded into continuous time using the technique discussed in Section 1.2.2. In the networks language, families correspond to nodes and the size of a family corresponds to the degree of the node, or alternatively the number of half-edges connected to the node. The size of the population corresponds to the number of half-edges in the network. At the nth "birth event" a new node is introduced into the network and connects to a randomly chosen existing node. This is equivalent to a selectant being born, since an old node increases its degree by one, and also a mutant being born, since the incoming node has degree one. Note that the RBP representation only keeps track of vertices and their degrees, but not of the actual edges (see [26]).

Condensation

Unlike the Barabási–Albert model where the degree of the network is correlated with its age, in the Bianconi–Barabási network young nodes have a chance to compete, if they get a relatively high fitness value. This leads to *self-organised criticality*, when at a large finite time old and young nodes compete on the same level. Furthermore, it is possible that a positive proportion of individuals has fitnesses converging to the maximal possible value, a phenomenon referred to as *condensation*, by comparison with Bose–Einstein condensation (see [15]). Presence or absence of condensation is determined by the existence of the Malthusian growth rate (see [26, Theorem 2.1]).

Condensation can be of two types: *macroscopic* and *non-extensive* (see [26]). Macroscopic condensation refers to the case when the size of the largest family (maximum degree in the network) occupies a non-zero proportion of the links. Non-extensive condensation is when the so-called condensate is formed by a collection of families. In [26, Theorem 2.4] Dereich, Mailler and Mörters prove that for finite, regularly varying fitness distributions macroscopic condensation does not occur in the RBPs, which they call "The winner does not take it all". This contradicts conjectures made in [15], and highlights how subtle the condensation phenomenon is.

Limit characteristics of the largest family

Apart from studying the phenomenon of condensation, Dereich, Mailler and Mörters [26] analyse the asymptotic limits of the largest family in the population. They "zoom" into a window from which the largest family can originate and find a Poisson limit for particles contained in it [26, Theorem2.2]. This gives the limits for the size, fitness and time of formation of the largest family [26, Corollary 2.3]. The analysis in [26] is conducted for bounded, regularly varying fitness distributions, which covers the distributions lying in the maximum domain of attraction (MDA) of Weibull distributions. In this thesis we extend this analysis to a more general growth processes framework and study the asymptotic characteristics of the largest family for another class of bounded fitness distributions.

1.3 Competing growth processes (CGPs)

We build upon [26], and generalise the RBPs to what we call *competing growth processes* (or CGPs for short). This generalisation allows an arbitrary number of offspring born at a single birth event. Apart from the preferential attachment tree of Bianconi and Barabási and branching processes with selection and mutation mentioned above, our framework allows us to include into our analysis the *preferential attachment network with fitness of Dereich* (introduced in [27]) and variations of Chinese restaurants (see for example [13]), for which we derive a surprising result on the relative size of the two largest occupied tables. We discuss each of these examples in detail in Chapter 4.

1.3.1 Motivational calculation

In this thesis we investigate the asymptotic properties of the largest family in a sequence of growing families, which have different birth times and different exponential growth rates. The growth rates are sampled from an i.i.d. sequence $F_1, F_2, \dots$ of bounded random variables, while the birth times $\tau_1, \tau_2, \dots$ may be random and can depend in quite a general fashion on the growth processes. In the most interesting cases the birth times are themselves arising from an exponentially growing process so that the largest family at time t arises in competition of the few families born early, which have a longer time to grow, and the many families born late, among which the occurrence of a higher birth rate is more probable. The situation we investigate arises for example in various dynamic network models, where the families are nodes and their size is the degree, or in variants of the Chinese restaurant processes, where the families are tables and their sizes is the number of occupants. We give a flavour of the problem by a calculation based on the simplest nontrivial scenario.

For this purpose let the birth time of the nth family be $\tau_n = \frac{1}{\lambda}\log n$ and its size at time t be

$$Z_n(t) = \begin{cases} \lfloor \mathrm{e}^{(t-\tau_n)F_n} \rfloor & \text{if } \tau_n < t, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose μ is the law of F_n on the interval $(0,1]$ and let $1 \ll T(t) \ll t$. Then

$$\begin{aligned}
\mathbb{P}\Big(\mathrm{e}^{-(t-T(t))} \max_n Z_n(t) \le \mathrm{e}^x\Big) &= \mathbb{P}((t-\tau_n)F_n \le (t-T(t)) + x \ \forall n \colon \tau_n \le t) \\
&= \prod_{\tau_n \le T(t)-x} \mathbb{P}\Big(F_n \le \frac{t-T(t)+x}{t-\tau_n}\Big) \\
&= \exp\Big(\sum_{n \le \mathrm{e}^{\lambda(T(t)-x)}} \log\Big(1 - \mu\Big(\big(\tfrac{t-T(t)+x}{t-\tau_n}, 1\big]\Big)\Big)\Big) \\
&= \exp\Big(-(1+o(1)) \sum_{n \le \mathrm{e}^{\lambda(T(t)-x)}} \mu\Big(\big(\tfrac{t-T(t)+x}{t-\tau_n}, 1\big]\Big)\Big).
\end{aligned}$$

The task is now to choose $T(t)$ such that, as $t \uparrow \infty$,

$$\sum_{n \le \mathrm{e}^{\lambda(T(t)-x)}} \mu\Big(\big(\tfrac{t-T(t)+x}{t-\tau_n}, 1\big]\Big) \longrightarrow \phi(x),$$

for some nondegenerate function ϕ. The solution depends on the tail of μ at one.

Supposing for example that $\mu((1-x,1]) \sim x^\alpha$ as $x \downarrow 0$, for some index $\alpha > 0$, we get

$$\sum_{n \leq e^{\lambda(T(t)-x)}} \mu\Big(\big(\tfrac{t-T(t)+x}{t-\tau_n}, 1\big]\Big) \sim \frac{1}{t^\alpha} \sum_{n \leq e^{\lambda(T(t)-x)}} \big(T(t) - \tau_n - x\big)^\alpha.$$

Letting $T(t) = \frac{\alpha}{\lambda} \log t$ this is equivalent to

$$\frac{1}{t^\alpha} \int_0^{t^\alpha e^{-\lambda x}} \Big(-\tfrac{1}{\lambda} \log\big(\tfrac{n}{t^\alpha}\big) - x\Big)^\alpha \, dn = e^{-\lambda x} \int_0^\infty \lambda e^{-\lambda u} u^\alpha \, du = e^{-\lambda x} \lambda^{-\alpha} \Gamma(\alpha+1),$$

using the substitution $u = -\frac{1}{\lambda} \log\big(\frac{n}{t^\alpha}\big) - x$. Hence we have that

$$e^{-t} \big(\tfrac{(\lambda t)^\alpha}{\Gamma(\alpha+1)}\big)^{\frac{1}{\lambda}} \max_n Z_n(t) \Longrightarrow \Phi_\lambda,$$

where Φ_λ is the Fréchet distribution with parameter λ.

This result, and further asymptotic results on the birth time and fitness of the largest family, can be generalised to a framework where

- μ is in the maximum domain of attraction of the *Weibull distribution* of extreme value theory,
- the growth processes $(Z_n(\tau_n + s) \colon s \geq 0)$ are asymptotically independent random processes with growth rates given as γF_n, for some $\gamma > 0$,
- the birth times τ_n are themselves random and may depend on the growth processes.

Generalising the above calculation to such a setup requires, of course, more sophisticated methods. Our approach is to describe the state of a family at time t as a point in the space $(-\infty, \infty) \times (-\infty, \infty) \times (0, \infty)$, where the first coordinate corresponds to its birth time, the second to its fitness and the third to its size at time t. Introducing a t-dependent scaling of the three coordinates (so that the focus is on a carefully chosen *window*) and letting $t \to \infty$ we obtain a limiting point process, see Theorem 1. In this limiting process the point with the maximal third coordinate identifies the largest family, allowing to read off limit theorems for its size, fitness and birth time, see Corollary 2. A similar result in a different framework is contained in the paper [26].

1.3.2 Main results for two classes of fitness distributions

Our results describe size, fitness and birth time of the largest family in CGPs at a large time t. From an applications perspective, these correspond to the properties of the vertex with a maximal degree in a preferential attachment network, of the largest family in a population with selection and mutation and the largest table in a Chinese restaurant process. These properties are analysed in [26] under the framework of RBPs for a class of bounded fitness distributions μ with regularly varying tail, or equivalently characterised as distributions lying in the maximum domain of attraction of the *Weibull distribution*, MDA(Weibull). We generalise the analysis to our setup in Theorem 1 and Corollary 2 (proofs are in Chapter 3). Our *main results*, Theorem 3 and Corollary 4, provide corresponding results for the case that μ is in the maximum domain of attraction of the *Gumbel distribution*, MDA(Gumbel), defined and illustrated in Section 1.4.1. This case is considerably more difficult because the technique of [26] cannot be applied directly (see Chapter 2).

The reason for this is that the "window" in which one has to search for the largest family is bigger, having unbounded width in the first component. Therefore for a limit theorem the first component requires scaling, and hence the scaling of the second component depends not only on t but also on n, the birth rank of the family. Using some additional regularity properties of the fitness distribution μ allows to make the scaling of the third component independent of n, so that we can still achieve a powerful Poisson limit theorem (Theorem 3) as well as convergence of the scaled family size to a Fréchet distribution and of the standardised birth time to a Gaussian distribution (Corollary 4).

This way we get results for bounded fitness distributions lying in MDA(Weibull) and MDA(Gumbel). Functions lying in MDA(Fréchet) have unbounded support (see for example [34, Chapter 3.3.1]); they are more difficult to study and constitute an open area of research (see Section 1.3.3 for a discussion). Taken together, our results give an essentially complete picture for the behaviour of the largest family for fitness distributions μ with bounded support.

1.3.3 Open questions

Many interesting questions arising from the applications of the CGPs are yet to be answered. The examples we discuss inspire interest in different aspects of the model. A classical question asked about networks concerns their *degree distribution*. Since the degree distribution for RBPs is scale-free it is reasonable to expect that CGPs also have this property; we leave this as an open problem. For preferential attachment networks with fitness one could also ask about their *joint degree and fitness distribution*. For RBPs it has been shown in [26] and for a variant of the Bianconi–Barabási network in [29]. Though it is probably possible to apply similar techniques for the analysis of the joint degree and fitness distribution of CGPs, we have not attempted this ourselves.

Other obvious questions from the networks perspective concern growth dynamics of the system for the case with *unbounded fitness* distributions, and for the case with *condensation*. Unbounded fitnesses with light tails could lead to the empirical fitness distribution (defined as $\Xi_t := \frac{1}{N(t)} \sum_{n=1}^{M(t)} Z_n(t)\delta_{F_n}$), splitting into two parts: the so-called *bulk*, with asymptotic shape of μ and the *travelling wave*, a part of the mass going to infinity. The latter is particularly difficult to study, with possible questions regarding the speed of travel, the spread and the asymptotic shape of the wave. Fitnesses with unbounded support and heavy tails could lead to *explosion*; one question that could be asked is characterising the network's behaviour just before explosion occurs. Condensation is little understood and it is an exciting open problem, with many papers being published on the topic. See for example [40] for a recent work on the extensive condensation in preferential attachment networks. The condensation case is challenging due to the absence of a Malthusian growth parameter, which means that there is very little control over the growth of the system, and so new techniques need to be developed for its analysis.

One can think of three natural generalisations for populations with selection and mutation, namely, incorporating *dependency of the mutant's fitness* on the parent, allowing offspring to have *multiple parents* and making individuals mortal by introducing *death rates*. These modifications seem to be easy to formulate, but make the analysis more delicate: the first two lead to the loss of independence between individuals whereas the third leads to extra randomness.

A classical question in the literature on Chinese restaurant processes concerns the *ratio between the sizes of tables* (see for example [13]). Here we are able to calculate the ratio between the first and the second largest tables for our variant of the process (see Section 4.3). It would be interesting to analyse the ratios for the consecutive tables sorted by growth, i.e. second and third largest tables, third and fourth, etc., and to check if the ratio has a universal property.

1.3.4 Structure of the thesis

The thesis is structured as follows. In Section 1.4 we give a rigorous definition of our model and assumptions and state the main results. Section 1.4.1 gives examples of fitness distributions to which our results apply. In Chapters 2 and 3 we prove *convergence to the Poisson limit* theorems and their corollaries for when the fitness distribution μ lies in the Gumbel and Weibull maximum domain of attraction, respectively. We begin with the Gumbel case, because it is more illustrative by requiring more sophisticated methods. In Section 2.1.1 we state some general results about branching processes, needed for our proofs. In Sections 2.1.2 and 2.1.3 we prove local convergence for a simpler point process. We prove the Poisson limit in Section 2.2 by considering contribution of young and old families, and families with small fitnesses. The proof for limits of family characteristics follows in Section 2.2.5. The contents of Chapter 2 are published in [56, Chapters 3, 4]. The proofs for the Weibull case in Chapter 3 follow a similar structure and use the methodology developed in Chapter 2. We prove local convergence for a simpler point process in Section 3.1, Poisson limit in Sections 3.2 and 3.3.1 and limits of family characteristics follow in Section 3.3.2.

Chapter 4 is devoted to the applications of CGPs and it is presented in [56, Chapters 2, 5]. In Sections 4.1–4.3 we define each of the examples in turn and explain the implications of our results for each of these models. We then prove that these examples satisfy our assumptions and therefore fall under our framework (Section 4.4). Chapters are reasonably independent and therefore can be read in any order. However reading them in a linear order benefits from the logical structure of the thesis.

1.4 Definition of the model and main results

In this section we define the model, state our assumptions and main results. We give examples of fitness distributions that satisfy these assumptions and examples covered by our framework.

Let μ be a probability distribution on $(0, 1)$ and

- $(F_n)_{n\geq 1}$ be i.i.d. μ-distributed random variables;
- $(\tau_n)_{n\geq 1}$ be a non-decreasing sequence of positive random variables with $\tau_1 = 0$;
- $Z_n(t) = X_n(F_n(t - \tau_n))$ for a family $(X_n(t) : t \geq 0)_{n\geq 1}$ of non-decreasing integer valued processes.

Define $M(t) := \max\{n : \tau_n \leq t\}$ and $N(t) := \sum_{n=1}^{M(t)} Z_n(t)$. We view this as a population of immortal individuals and we refer to $Z_n(t)$ as the size of the nth family, $M(t)$ the number of families in the system and $N(t)$ the total size of the population respectively, at time t. From this perspective τ_n represents the foundation time of the nth family. Furthermore, we see F_n as a fitness parameter of the nth family, determining the rate at which new offspring are born into it.

In this research we aim at proving convergence results for the maximal family in the population. For this we require the following assumptions on the growth processes and fitness distribution.

Assumption (A.1) (Families' foundation times)**.** *There exists $\lambda > 0$ such that for all $n \in \mathbb{N}$*

$$\tau_n = \tau_n^* + T + \varepsilon_n,$$

where $\tau_n^* := \frac{1}{\lambda}\log n$, T is a finite random variable, and $\varepsilon_n \to 0$ almost surely as $n \to \infty$.

Note. *We require the almost sure convergence of ε_n to 0 for the proofs of Lemmas 13 and 23 below.*

Assumption (A.2) (Growth processes)**.** *There exist $\gamma > 0$ and an i.i.d. sequence of processes $((Y_n(t) : t \geq 0))_{n\geq 1}$ independent of $(F_n)_{n\geq 1}$, such that*

$$\Delta_n(t) := \sup_{u\geq t} \mathrm{e}^{-\gamma u}\Big|X_n(u) - Y_n(u)\Big|$$

satisfies for all $\varepsilon > 0$, and $\kappa > 0$

$$\sup_{n\in I_\kappa(t)} \mathbb{P}(\Delta_n(t) \geq \varepsilon\big|\,(F_m)_{m\in\mathbb{N}}) \to 0, \qquad \textit{in probability as } t \to \infty, \tag{1.5}$$

where $I_\kappa(t)$ is a collection of indices specified below in dependence on the fitness distribution μ.

Assumption (A.3) (Growth rate)**.** *There exists an integrable random variable ξ with density ν defined on $[0, \infty)$, such that*

$$\mathrm{e}^{-\gamma t}Y_1(t) \longrightarrow \xi, \qquad \textit{almost surely as } t \to \infty.$$

Assumption (A.4) (Concentration of growth)**.** *There exist $c_0, \eta > 0$ such that, for $n \in \mathbb{N}$, we have*

$$\mathbb{P}(\max_{u\geq 0} X_n(u)\mathrm{e}^{-\gamma u} \geq x\big|\,(F_m)_{m\in\mathbb{N}}) \leq c_0\mathrm{e}^{-\eta x}, \qquad \textit{for all } x \geq 0.$$

Beyond these four assumptions on the growth processes we need assumptions on the fitness distribution μ. We discuss two different possible classes of fitness distributions μ. The first class, corresponds to μ being in the maximum domain of attraction of the Gumbel distribution. We make the following assumptions.

Assumption (A.5) (μ in the maximum domain of attraction of the Gumbel distribution)**.**
The function $m(x) = -\log \mu(x, 1)$ is twice differentiable and satisfies

(A5.1) $m'(x) > 0$ *and* $m''(x) > 0$ *for all* $x \in [0, 1]$*;*

(A5.2) $\lim_{x\uparrow 1} \frac{m''(x)}{(m'(x))^2} = 0$*;*

(A5.3) $\exists \varkappa > 0$ *such that* $\lim_{x\uparrow 1} \frac{m''(x)m(x)x}{(m'(x))^2} = \varkappa$*;*

(A5.4) $\lim_{x\uparrow 1} \frac{m(x)}{m'(x)} = 0$.

Note. *Assumption (A.5) is sufficient for μ to be in the maximum domain of attraction of the Gumbel distribution, and contains the most important cases, but it is not formally necessary. We discuss this further in Section 1.4.1.*

Under Assumption (A.5) we define σ_t as the unique solution of

$$(\log g)'(\lambda\sigma_t) = \frac{1}{\lambda(t-\sigma_t)}, \tag{1.6}$$

where $g(x) = m^{-1}(x)$, see Lemma 7 for a proof of existence and uniqueness of σ_t. We then define the collection of indices in (A.2) as

$$I_\kappa(t) := \{n : \tfrac{|\tau_n^* - \sigma_t|}{\sqrt{\sigma_t}} \le \kappa\}, \quad \text{for } \kappa > 0. \tag{1.7}$$

The other class of distributions μ we consider is the maximum domain of attraction of the Weibull distribution class. The proofs can be found in Chapter 3.

Assumption (B.5) (μ in the maximum domain of attraction of the Weibull distribution)**.** *The fitness distribution μ has a regularly varying tail at $x = 1$, meaning that there exists $\alpha > 0$ and a slowly varying function ℓ with $\mu(1-\varepsilon, 1) = \varepsilon^\alpha \ell(\varepsilon)$.*

We set

$$\sigma_t := \tau_{n(t)}, \quad \text{where} \quad n(t) = \left\lceil \frac{1}{\mu(1-t^{-1}, 1)} \right\rceil \tag{1.8}$$

and use this to define

$$I_\kappa(t) := \{n : |\tau_n^* - \sigma_t| \le 2|T| + \kappa\}, \quad \text{for } \kappa > 0, \tag{1.9}$$

for use in Assumption (A.2). Assumption (B.5) implies that $n(t) = \lceil \frac{t^\alpha}{\ell(t^{-1})} \rceil$ and so $\log n(t) \sim \alpha \log t$. Using this we can write

$$\tau_{n(t)} = \frac{1}{\lambda}\log n(t) + T + \varepsilon_{n(t)} = \frac{\alpha}{\lambda}\log t + T - \frac{1}{\lambda}\log \ell(t^{-1}) + o(1),$$

as $t \to \infty$, by Assumption (A.1).

We now state our results, first in the easier case of μ satisfying Assumption (B.5). For all $t \ge 0$, we define the point process

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta\big(\tau_n - \sigma_t, t(1-F_n), \mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t)\big), \tag{1.10}$$

on $(-\infty,\infty) \times (0,\infty) \times (0,\infty)$, where $\delta(x)$ is the Dirac mass at x. We will look at the limits of Γ_t, strengthening the result considerably by partially compactifying the underlying space.

Theorem 1 (Poisson limit)**.** *Under assumptions (A.1) to (A.4) and (B.5) the point process $(\Gamma_t)_{t\ge 0}$ converges vaguely*[4] *in distribution on the space $[-\infty,\infty] \times [0,\infty] \times (0,\infty]$ to the Poisson point process with intensity measure*

$$\mathrm{d}\zeta(s,f,z) = \alpha f^{\alpha-1} \lambda \mathrm{e}^{\lambda s} \mathrm{e}^{\gamma(s+f)} \nu(z\mathrm{e}^{\gamma(s+f)})\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z,$$

where ν is as in (A.3).

Observe that the compactification of the intervals in Theorem 1 ensures that the point with the largest z-component in the Poisson process corresponds asymptotically to the family of maximal size. Theorem 1 therefore implies the following distributional limits (denoted by $\Rightarrow$) for the size, fitness and the foundation time of the largest family.

Corollary 2 (Limits of family characteristics)**.** *Under the same assumptions as in Theorem 1, we have*

(i) Asymptotically, as $t \to \infty$,

$$\mathrm{e}^{-\gamma t + \frac{\gamma\alpha}{\lambda}\log t + \gamma T} \max_{n\in\mathbb{N}} Z_n(t) \Rightarrow W,$$

where W is Fréchet distributed with shape parameter λ/γ and scale parameter s, where

$$s^{\frac{\lambda}{\gamma}} = \Gamma(\alpha+1)\lambda^{-\alpha}\int_0^\infty \nu(w) w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w.$$

(ii) Denoting by $V(t)$ the fitness of the family of maximal size at time t, as $t \to \infty$, we have

$$t(1 - V(t)) \Rightarrow V,$$

where V is Gamma distributed with shape parameter α and scale parameter $\frac{1}{\lambda}$.

(iii) Denoting by $S(t)$ the birth time of the family of maximal size at time t, as $t \to \infty$, we have

$$S(t) - \sigma_t \Rightarrow U,$$

where U is a real valued random variable.

Note. *Substituting $\tau_n = \frac{1}{\lambda}\log n + T + \varepsilon_n$ into Corollary 2(i) one can show that for discrete time it translates to $\max_{m\geq n} Z_m(\tau_n) = n^{\gamma/\lambda + o_n(1)}$. For $\gamma = 1$ this result corresponds to the largest degree in the Bianconi–Barabási tree with fitness, which has a power-law degree distribution with rate $\lambda + 1$ (see Section 1.2.2) and so our result is consistent with Equation* (1.2).

Theorem 1 and Corollary 2 are proved in Chapter 3. A similar result in a different setup has been shown in [26] using a different approach.

To now state *our main results* we look at fitness distributions satisfying Assumption (A.5). For all $t \geq 0$, we define

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta\Big(\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T} Z_n(t)\Big), \tag{1.11}$$

[4] A sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ on a topological space $\mathbb{X}$ converges *vaguely* to μ iff $\int f \mathrm{d}\mu_n \to \int f\mathrm{d}\mu$, as $n \to \infty$, for all continuous functions $f\colon \mathbb{X} \to \mathbb{R}$ with compact support.

where $\delta(x)$ is the Dirac mass at x, and $a_1 := \frac{\gamma}{2\lambda}$.

Theorem 3 (Poisson limit)**.** *Under assumptions (A.1) to (A.5) the point process $(\Gamma_t)_{t\geq 0}$ converges vaguely in distribution on the space $[-\infty,\infty]\times[-\infty,\infty]\times(0,\infty]$ to the Poisson point process with intensity measure*

$$\mathrm{d}\zeta(s,\ f,\ z) = \lambda \mathrm{e}^{-f}\mathrm{e}^{s^2 a_2 - f a_3}\nu(z\mathrm{e}^{s^2 a_2 - f a_3})\,\mathrm{d}s\,\mathrm{d}f\,\mathrm{d}z,$$

where $a_2 := \frac{\gamma}{2}\varkappa$, $a_3 := \frac{\gamma}{\lambda}$ and ν is as in (A.3).

Note. *The existence of a density for the random variable ξ is assumed in (A.3) for convenience. For example, Theorems 1 and 3 continue to hold if $\nu = \delta_1$ as in our motivating example.*

The technical difference between Theorems 1 and 3 is that in the latter the first (birth time) coordinate needs to be scaled. As a result the scaling of the second (fitness) component depends on the birth rank n of the family as well as on the observation time t. Therefore we cannot derive a general scaling limit for the fitness of the largest family as in Corollary 2. Results for the birth time and size of this family, however, are still possible.

Corollary 4 (Limits of family characteristics)**.** *Under the same assumptions as in Theorem 3, we have*

(i) Asymptotically as $t\to\infty$,

$$\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T}\max_{n\in\mathbb{N}} Z_n(t) \Rightarrow W,$$

where W is Fréchet distributed with shape parameter λ/γ and scale parameter s where

$$s^{\frac{\lambda}{\gamma}} = \sqrt{\tfrac{2\pi\lambda}{\varkappa}}\int_0^\infty \nu(w)w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w.$$

(ii) Denoting by $S(t)$ the birth time of the family of maximal size at time t, as $t\to\infty$, we have

$$\frac{S(t)-\sigma_t}{\sqrt{\sigma_t}} \Rightarrow U,$$

where U is normally-distributed with mean 0 and variance $\frac{1}{\lambda\varkappa}$.

Note. *Observe that irrespective of whether μ is in the maximum domain of attraction of the Weibull or Gumbel distribution, the size of the largest family scaled by a determinstic function of time and the random factor $\mathrm{e}^{\gamma T}$ converges to a Fréchet distribution.*

1.4.1 Examples of fitness distributions

The five following functions $m(x) = -\log\mu(x,1)$ satisfy Assumption (A.5):

(1) $m(x) = (1-x)^{-\varrho} - 1$, where $\varrho > 0$;

(2) $m(x) = \mathrm{e}^{\frac{1}{1-x}} - \mathrm{e}$;

(3) $m(x) = \frac{x}{1-x}$;

(4) $m(x) = \mathrm{e}^{\frac{1}{\sqrt{1-x}}} - \mathrm{e}$;

(5) $m(x) = \tan\left(\frac{\pi x}{2}\right)$.

In the following lemma, we check that function in example (1) satisfies the conditions of Assumption (A.5), but we omit calculations for examples (2)–(5), since they are straightforward.

Lemma 5. *The function $m(x) = (1-x)^{-\varrho}-1$, where $\varrho > 0$, satisfies Assumption (A.5).*

Proof. First of all, we have $m(0) = 0$ and $m(1) = \infty$, so that $\mu(x,1) = \mathrm{e}^{-m(x)}$ is a probability measure. We now consider each condition (A5.1) to (A5.4) in order. We have $m'(x) = \varrho(1-x)^{-(\varrho+1)} > 0$ and $m''(x) = \varrho(\varrho+1)(1-x)^{-(\varrho+2)} > 0$, for all $x \in [0,1]$, which implies (A5.1). Also, we have that

$$\lim_{x\uparrow 1} \frac{m''(x)}{(m'(x))^2} = \lim_{x\uparrow 1} \frac{(\varrho+1)(1-x)^{\varrho}}{\varrho} = 0,$$

which implies (A5.2) and

$$\lim_{x\uparrow 1} \frac{m''(x)m(x)x}{(m'(x))^2} = \lim_{x\uparrow 1} \frac{(\varrho+1)x(1-(1-x)^{\varrho})}{\varrho} = \frac{\varrho+1}{\varrho} > 0, \tag{1.12}$$

which implies (A5.3) and $\varkappa = \frac{\varrho+1}{\varrho}$. Finally, we get

$$\lim_{x\uparrow 1} \frac{m(x)}{m'(x)} = \lim_{x\uparrow 1} \frac{(1-x)^{-\varrho}-1}{\varrho(1-x)^{-(\varrho+1)}} = 0,$$

which implies (A5.4). $\square$

Assumptions (A5.1) and (A5.2) imply that the fitness distribution μ lies in the maximum domain of attraction of the Gumbel distribution, see [34, Chapter 3.3.3]. More precisely, let $(F_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with common distribution μ, then

$$\frac{\max_{i\leq n} F_i - g(\log n)}{g'(\log n)} \to V, \quad \text{in distribution as } n \to \infty, \tag{1.13}$$

where $V \sim \Lambda$, i.e. $\mathbb{P}(V \leq x) = \exp\{-\mathrm{e}^{-x}\}$, for all $x \in \mathbb{R}$ (for more details see [68, Chapter 0.3]).

Although most of the natural examples satisfy Assumptions (A5.3) and (A5.4), some probability distributions in the maximum domain of attraction of the Gumbel distribution do not fall into our framework, for example

(6) $m(x) = \log\left(\frac{\mathrm{e}}{1-x}\right)\log\log\left(\frac{\mathrm{e}}{1-x}\right)$.

Lemma 6. *If $m(x) = \log\left(\frac{\mathrm{e}}{1-x}\right)\log\log\left(\frac{\mathrm{e}}{1-x}\right)$, then the fitness distribution $\mu(x,1) = \mathrm{e}^{-m(x)}$ lies in MDA(Gumbel), but Assumptions (A5.3) and (A5.4) are not satisfied.*

Proof. First of all $m(0) = 0$ and $m(1) = \infty$, implying that $\mu(x,1) = \mathrm{e}^{-m(x)}$ is a probability measure. For all $x \in [0,1]$, we have

$$m'(x) = \frac{\log(\log\frac{\mathrm{e}}{1-x})}{1-x} > 0,$$

and expressing $\log(\log\frac{\mathrm{e}}{1-x}) = \log(1-\log(1-x))$, we can see that

$$m''(x) = \frac{\frac{1}{1-\log(1-x)} + \log(1-\log(1-x))}{(1-x)^2} > 0,$$

since $-\log(1-x) > 0$. Furthermore, we have

$$\lim_{x\uparrow 1}\frac{m''(x)}{(m'(x))^2} = \lim_{x\uparrow 1}\frac{\frac{1}{\log\frac{\mathrm{e}}{1-x}} + \log\log\frac{\mathrm{e}}{1-x} + 1}{(\log\log\frac{\mathrm{e}}{1-x}+1)^2} = 0,$$

so Assumptions (A5.1) and (A5.2) are satisfied, and μ lies in MDA(Gumbel). However, we have

$$\lim_{x\uparrow 1}\frac{m''(x)m(x)x}{(m'(x))^2} = \lim_{x\uparrow 1}\left(\log\frac{\mathrm{e}}{1-x}\log\log\frac{\mathrm{e}}{1-x}x\right)\left(\frac{\frac{1}{\log\frac{\mathrm{e}}{1-x}} + \log\log\frac{\mathrm{e}}{1-x} + 1}{\log\log\frac{\mathrm{e}}{1-x}}\right) = \infty,$$

and

$$\lim_{x\uparrow 1}\frac{m(x)}{m'(x)} = \lim_{x\uparrow 1}\frac{\log\frac{\mathrm{e}}{1-x}}{1-x} = \infty,$$

which means that Assumptions (A5.3) and (A5.4) are both not satisfied. $\square$

Chapter 2

μ in MDA(Gumbel)

In this Chapter we prove Theorem 3 and Corollary 4. In Section 2.1 we look at the Poisson limit theorem given in Theorem 3, but first in a space without compactifications. After some preparations we prove in Section 2.1.2 a basic form of the limit theorem, see Proposition 9. This is derived from an approximation which corresponds to a classical Poisson convergence result for extremes in the first two components and an independent third component. In Section 2.1.3 a further approximation turns the basic form into the original form of the Poisson limit theorem, the crucial difference being that the scaling of the third component becomes independent of the birth rank n of the family. Section 2.2 is devoted to the compactification of the space, effectively showing that the points supressed by the scalings do not provide the largest families. These points are either born too late (Section 2.2.1) or not fit enough (Section 2.2.2). In Section 2.2.3 we show that there are no points outside our scaling window that are competitive in age and fitness. The proof of Theorem 3 is completed in Section 2.2.4 and the proof of Corollary 4, which crucially uses the compactification, in Section 2.2.5.

2.1 Local convergence of point processes

In this section we prove convergence result for the point processes (Γ_t) and its approximations in a space without compactification. The strengthening of the results by compactification will follow in the next section. We begin by noting some preliminary results on the fitness distribution.

2.1.1 Preliminaries on the fitness distribution

First of all we show the existence and uniqueness of σ_t as defined in Equation (1.6).

Lemma 7. *For all t large enough, there exists a unique $\sigma_t \in [0, t]$, such that*

$$(\log g)'(\lambda \sigma_t) = \frac{1}{\lambda(t - \sigma_t)}.$$

Furthermore, we have $\sigma_t \to \infty$ and $\frac{\sigma_t}{t} \to 0$ as $t \to \infty$.

Proof. Let
$$F(x) := (\log g)'(x) - \frac{1}{\lambda t - x},$$
so F is continuous on $(0, \lambda t)$. Since $g(x) = m^{-1}(x)$ and $g(0) = 0$, we have
$$\lim_{x \downarrow 0} (\log g)'(x) = \lim_{x \downarrow 0} \frac{g'(x)}{g(x)} = \infty.$$
Therefore we get
$$\lim_{x \downarrow 0} F(x) = \infty,$$
and
$$\lim_{x \uparrow \lambda t} F(x) = -\infty.$$
Hence by continuity of F, there exists $x \in (0, \lambda t)$ such that $F(x) = 0$. Furthermore such x is unique because
$$F'(x) = \frac{g''(x)}{g(x)} - \left(\frac{g'(x)}{g(x)}\right)^2 - \left(\frac{1}{\lambda t - x}\right)^2 < 0 \quad \text{for all } x \in (0, \lambda t),$$
since $g''(x) = -\frac{m''(g(x))}{(m'(g(x)))^3} < 0$ by Assumption (A5.1). Setting $\sigma_t = \frac{1}{\lambda} x$ proves existence and uniqueness as required, moreover σ_t is increasing in t.

It remains to show that $\sigma_t \to \infty$ as $t \to \infty$. If σ_t was bounded, we would have $\frac{1}{\lambda t - \lambda \sigma_t} \to 0$ as $t \to \infty$. This implies that $(\log g)'(\lambda \sigma_t) = \frac{g'(\lambda \sigma_t)}{g(\lambda \sigma_t)} \to 0$ and hence $g'(\lambda \sigma_t) \to 0$. This means $\frac{1}{m'(g(\lambda \sigma_t))} \to 0$, i.e. $m'(g(\lambda \sigma_t)) \to \infty$. From Assumption (A5.4), we know that $m'(x) \uparrow \infty$ as $x \uparrow 1$ and therefore $g(\lambda \sigma_t) \to 1$ and hence $\sigma_t \to \infty$ as required.

Finally we show that $\frac{\sigma_t}{t} \to 0$. By definition of σ_t, we have $t = \sigma_t + \frac{g(\lambda \sigma_t)}{\lambda g'(\lambda \sigma_t)}$, so we can write
$$\lim_{t \to \infty} \frac{\sigma_t}{t} = \lim_{t \to \infty} \frac{\sigma_t}{\sigma_t + \frac{g(\lambda \sigma_t)}{\lambda g'(\lambda \sigma_t)}} = \lim_{t \to \infty} \frac{1}{1 + \frac{g(\lambda \sigma_t)}{\lambda \sigma_t g'(\lambda \sigma_t)}}.$$
As $\sigma_t \to \infty$ as have $g(\lambda \sigma_t) \to 1$ as $t \to \infty$ we get
$$\lim_{t \to \infty} \frac{\sigma_t}{t} = \lim_{x \uparrow 1} \frac{1}{1 + \frac{x}{\frac{m(x)}{m'(x)}}} = 0,$$
since $\lim_{x \uparrow 1} \frac{m(x)}{m'(x)} = 0$ by Assumption (A5.4). □

From Lemma 7, it follows that $\lambda t \sim \frac{g(\lambda \sigma_t)}{g'(\lambda \sigma_t)}$ as $t \to \infty$. As $g(\lambda \sigma_t) \to 1$, we get
$$g'(\lambda \sigma_t) \sim \frac{1}{\lambda t}. \tag{2.1}$$

Lemma 8. *We have*

$$\lim_{t\to\infty} \sigma_t t g''(\lambda\sigma_t) = -\varkappa\lambda^{-2}, \tag{2.2}$$

where $\varkappa$ is defined in Assumption (A5.3), and

$$\lim_{t\to\infty} \sigma_t g'(\lambda\sigma_t) = 0. \tag{2.3}$$

Proof. Recall that $\sigma_t = \frac{1}{\lambda} m(g(\lambda\sigma_t))$, $t \sim \frac{g(\lambda\sigma_t)}{\lambda g'(\lambda\sigma_t)}$ and

$$g''(x) = -\frac{m''(g(x))}{(m'(g(x)))^3} = -\frac{m''(g(x))g'(x)}{(m'(g(x)))^2},$$

since $m'(g(x)) = \frac{1}{g'(x)}$. Substituting these into (2.2) and substituting $x = g(\lambda\sigma_t)$, we get

$$\lim_{t\to\infty} \sigma_t t g''(\lambda\sigma_t) = \lim_{t\to\infty} -\frac{m(g(\lambda\sigma_t))g(\lambda\sigma_t)m''(g(\lambda\sigma_t))}{(\lambda m'(g(\lambda\sigma_t))^2} = \lim_{x\uparrow 1} -\frac{m''(x)m(x)x}{(\lambda m'(x))^2} = -\varkappa\lambda^{-2},$$

by Assumption (A5.3). Similarly, using $g'(\lambda\sigma_t) = \frac{1}{m'(g(\lambda\sigma_t))}$, we have

$$\lim_{t\to\infty} \sigma_t g'(\lambda\sigma_t) = \lim_{t\to\infty} \frac{m(g(\lambda\sigma_t))}{\lambda m'(g(\lambda\sigma_t))} = \lim_{x\uparrow 1} \frac{m(x)}{\lambda m'(x)} = 0,$$

by Assumption (A5.4). □

2.1.2 Convergence of a simpler point process

In this section we prove the following proposition, which gives a more basic form of the Poisson limit in a space without compactification.

Proposition 9. *We have vague convergence in distribution of the point process*

$$\Psi_t = \sum_{n=1}^{M(t)} \delta\Big(\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t)\Big)$$

to the Poisson point process with intensity

$$\zeta^*(\mathrm{d}s,\, \mathrm{d}f,\, \mathrm{d}z) = \lambda \mathrm{e}^{-f}\nu(z)\mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z,$$

on $(-\infty,\infty) \times (-\infty,\infty] \times [0,\infty]$.

We prove Proposition 9 in two steps:

(1) In Lemma 11 we approximate Ψ_t by the point process

$$\Psi_t^* = \sum_{n\in\mathbb{N}} \delta\Big(\frac{\frac{1}{\lambda}\log n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \xi_n\Big),$$

where we have replaced the rescaled family sizes $\mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t)$ by their limits, denoted ξ_n, and the birth times τ_n by the approximate birth times $\frac{1}{\lambda}\log n$, using Assumptions (A.3) and (A.1) respectively.

(2) In Lemma 10 we prove that Ψ_t^* converges to the Poisson point process with intensity ζ^*.

Lemma 10. *Under Assumption (A.5), in distribution when $t \to \infty$, $(\Psi_t^*)_{t\geq 0}$ converges vaguely on $(-\infty,\infty)\times(-\infty,\infty]\times[0,\infty]$ to the Poisson point process with intensity ζ^*.*

Proof. We apply Kallenberg's theorem, see [68, Proposition 3.22]. Since ζ^* is diffuse (atomless), to prove Lemma 10, it is enough to show that, for every precompact relatively open box $B \subset (-\infty,\infty)\times(-\infty,\infty]\times[0,\infty]$, we have

(a) $\mathbb{P}(\Psi_t^*(B) = 0) \to \exp(-\zeta^*(B))$, as $t \uparrow \infty$, and

(b) $\mathbb{E}[\Psi_t^*(B)] \to \zeta^*(B)$, as $t \uparrow \infty$.

It suffices to consider nonempty boxes B of the form $(s_0, s_1)\times(f_0, f_1)\times(z_0, z_1)$, where $s_0, s_1 \in (-\infty,\infty)$, $f_0, f_1 \in (-\infty,\infty]$, $z_0, z_1 \in [0,\infty]$, and $s_0 < s_1$, $f_0 < f_1$, $z_0 < z_1$. Note that

$$\zeta^*(B) = \lambda(s_1 - s_0)(\mathrm{e}^{-f_0} - \mathrm{e}^{-f_1}) \int_{z_0}^{z_1} \nu(x)\mathrm{d}x.$$

(a) Let

$$r(a) := \exp\left(\lambda(a\sqrt{\sigma_t} + \sigma_t)\right), \quad \text{for all } a,$$

and consider

$$\begin{aligned}\hat{\Psi}_t^* &= \sum_{n\in\mathbb{N}} \delta\Big(\frac{\frac{1}{\lambda}\log n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}\Big),\\ \hat{\zeta}^*(\mathrm{d}s,\ \mathrm{d}f) &= \lambda \mathrm{e}^{-f}\mathrm{d}s\,\mathrm{d}f.\end{aligned}$$

For $\hat{B} = (s_0, s_1)\times(f_0, f_1)$, we have $\hat{\zeta}^*(\hat{B}) = \lambda(s_1 - s_0)(\mathrm{e}^{-f_0} - \mathrm{e}^{-f_1})$. Denote, for $a > 0$,

$$\hat{f}_a(x) = g\big(\log(a\sqrt{\sigma_t})\big) + x g'\big(\log(a\sqrt{\sigma_t})\big).$$

Then we have

$$\begin{aligned}\mathbb{P}(\hat{\Psi}_t^*(\hat{B}) = 0) &= \prod_{n=r(s_0)}^{r(s_1)} \mathbb{P}\Big(\frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)} \notin (f_0, f_1)\Big)\\ &= \prod_{n=r(s_0)}^{r(s_1)} \Big[\mu\Big(0, \hat{f}_n(f_0)\Big) + \mu\Big(\hat{f}_n(f_1), 1\Big)\Big]\\ &= \prod_{n=r(s_0)}^{r(s_1)} \Big[1 - \mu\Big(\hat{f}_n(f_0), \hat{f}_n(f_1)\Big)\Big].\end{aligned}$$

Using the fact that $e^{-\mu(x_0,x_1)} = 1 - \mu(x_0,x_1) + o(\mu(x_0,x_1))$ when $x_0, x_1 \to 1$, we get that when $t \uparrow \infty$

$$\begin{aligned}\mathbb{P}(\hat{\Psi}_t^*(\hat{B}) = 0) &\sim \prod_{n=r(s_0)}^{r(s_1)} \mathrm{e}^{-\mu\left(\hat{f}_n(f_0), \hat{f}_n(f_1)\right)} \\ &\sim \prod_{n=r(s_0)}^{r(s_1)} \exp\Big\{ -\mu(\hat{f}_n(f_0), 1) + \mu(\hat{f}_n(f_1), 1)\Big\}.\end{aligned}$$

Recalling that $\mu(x,1) = \mathrm{e}^{-m(x)}$, we get the following

$$\begin{aligned}\mathbb{P}(\hat{\Psi}_t^*(\hat{B}) = 0) &\sim \exp\Big\{ \sum_{n=r(s_0)}^{r(s_1)} -\mathrm{e}^{-m\left(\hat{f}_n(f_0)\right)} + \mathrm{e}^{-m\left(\hat{f}_n(f_1)\right)} \Big\} \\ &\sim \exp\Big\{ -\int_{r(s_0)}^{r(s_1)} \mathrm{e}^{-m\left(\hat{f}_x(f_0)\right)} \mathrm{d}x + \int_{r(s_0)}^{r(s_1)} \mathrm{e}^{-m\left(\hat{f}_x(f_1)\right)} \mathrm{d}x \Big\}.\end{aligned}$$

We now evaluate the integrals in the exponent. For $i = 0, 1$ we have

$$\begin{aligned}\int_{r(s_0)}^{r(s_1)} \mathrm{e}^{-m\left(\hat{f}_x(f_i)\right)} \mathrm{d}x &= \int_{r(s_0)}^{r(s_1)} \exp\Big\{ -m\Big(g\big(\log(x\sqrt{\sigma_t})\big) + f_i g'\big(\log(x\sqrt{\sigma_t})\big)\Big)\Big\}\,\mathrm{d}x \\ &= \int_{s_0}^{s_1} \lambda\sqrt{\sigma_t} r(y) \exp\Big\{ -m\Big(g\Big(\log\big(r(y)\sqrt{\sigma_t}\big)\Big) + f_i g'\Big(\log\big(r(y)\sqrt{\sigma_t}\big)\Big)\Big)\Big\}\,\mathrm{d}y,\end{aligned}$$

by the change of variables, with $x = \mathrm{e}^{\lambda(y\sqrt{\sigma_t} + \sigma_t)} = r(y)$. By the mean value theorem, for each $i = 0, 1$, there exists a constant

$$c_3 \in \big[g\big(\log\big(r(y)\sqrt{\sigma_t}\big)\big), g\big(\log\big(r(y)\sqrt{\sigma_t}\big)\big) + f_0 g'\big(\log\big(r(y)\sqrt{\sigma_t}\big)\big)\big],$$

such that

$$\begin{aligned}&m\Big(g\big(\log\big(r(y)\sqrt{\sigma_t}\big)\big) + f_i g'\big(\log\big(r(y)\sqrt{\sigma_t}\big)\big)\Big) \\ &\quad = m\Big(g\big(\log(r(y)\sqrt{\sigma_t})\big)\Big) + f_i g'\big(\log(r(y)\sqrt{\sigma_t})\big) m'\Big(g\big(\log(r(y)\sqrt{\sigma_t})\big)\Big) \\ &\qquad + \frac{1}{2}\Big(f_i g'\big(\log(r(y)\sqrt{\sigma_t})\big)\Big)^2 m''(c_3).\end{aligned}$$

Recall that by definition, for all $x \in \mathbb{R}$, $m(g(x)) = x$ and $g'(x) = \frac{1}{m'(g(x))}$, so the integral simplifies to

$$\begin{aligned}\int_{r(s_0)}^{r(s_1)} \mathrm{e}^{-m\left(\hat{f}_x(f_i)\right)} \mathrm{d}x &= \int_{s_0}^{s_1} \lambda\, r(y)\sqrt{\sigma_t}\, \mathrm{e}^{-\log(r(y)\sqrt{\sigma_t}) - f_i - \frac{1}{2}\left(f_i g'(\log(r(y)\sqrt{\sigma_t}))\right)^2 m''(c_3)}\,\mathrm{d}y \\ &= \lambda \int_{s_0}^{s_1} \exp\Big\{ -f_i - \frac{f_i^2}{2} g'\big(\log(r(y)\sqrt{\sigma_t})\big)^2 m''(c_3)\Big\}\,\mathrm{d}y.\end{aligned}$$

Assumption (A5.2) implies that

$$g'\big(\log(r(y)\sqrt{\sigma_t})\big)^2 m''(c_3) = \frac{m''(c_3)}{m'\big(g(\log(r(y)\sqrt{\sigma_t}))\big)^2} \to 0,$$

as $t \to \infty$. By the dominated convergence theorem, as $t \to \infty$ we get

$$\int_{r(s_0)}^{r(s_1)} \mathrm{e}^{-m\left(\hat{f}_x(f_i)\right)} \mathrm{d}x \quad = \quad \lambda \int_{s_0}^{s_1} \mathrm{e}^{-f_i+o(1)}\, \mathrm{d}y = \lambda(s_1 - s_0)\mathrm{e}^{-f_i} + o(1).$$

Therefore, as $t \to \infty$ we get

$$\begin{aligned} \mathbb{P}(\hat{\Psi}_t^*(\hat{B}) = 0) \quad &\sim \quad \exp\Big\{ -\lambda(s_1 - s_0)\mathrm{e}^{-f_0} + \lambda(s_1 - s_0)\mathrm{e}^{-f_1} + o(1)\Big\} \\ &\to \quad \exp\Big\{ -\lambda(s_1 - s_0)\big(\mathrm{e}^{-f_0} - \mathrm{e}^{-f_1}\big)\Big\} = \exp\big\{ -\hat{\zeta}^*(\hat{B})\big\}. \end{aligned}$$

Using Kallenberg's theorem, we thus get that, in distribution when $t \to \infty$, $\hat{\Psi}_t^*$ converge vaguely on $(-\infty, +\infty) \times (-\infty, +\infty]$ to the Poisson point process of intensity $\hat{\zeta}^*$. By assumption, $(F_n, \xi_n)_{n\geq 1}$ is a sequence of i.i.d. random variables with each F_n being independent of ξ_n. We have $\mathbb{P}(\xi_n \in (z_0, z_1)) = \int_{z_0}^{z_1} \nu(x)\mathrm{d}x$, which completes the proof of (a).

(b) To calculate the limit of $\mathbb{E}[\Psi_t^*(B)]$ we apply similar asymptotic estimates as in part (a), and get

$$\begin{aligned} \mathbb{E}[\Psi_t^*(B)] \quad &= \quad \sum_{r(a_0)\leq n\leq r(a_1)} \mu\Big(\hat{f}_n(f_0), \hat{f}_n(f_1)\Big) \times \mathbb{P}(\xi_1 \in [z_0, z_1]) \\ &\sim \quad \int_{r(s_0)}^{r(s_1)} \mu\Big(\hat{f}_x(f_0), \hat{f}_x(f_1)\Big) \times \mathbb{P}(\xi_1 \in [z_0, z_1])\, \mathrm{d}x \\ &\sim \quad \lambda(s_1 - s_0)\big(\mathrm{e}^{-f_0} - \mathrm{e}^{-f_1}\big) \int_{z_0}^{z_1} \nu(x)\, \mathrm{d}x \to \zeta^*(B), \quad \text{as } t \to \infty. \end{aligned}$$

□

Lemma 11. *For all Lipschitz continuous, compactly supported functions* $f : (-\infty, \infty) \times (-\infty, \infty] \times [0, \infty] \to \mathbb{R}$,

$$\left| \int f \mathrm{d}\Psi_t^* - \int f \mathrm{d}\Psi_t \right| \to 0 \text{ in probability, as } t \uparrow \infty.$$

Proof. Let f be a Lipschitz continuous function supported on $K = [-a, a] \times [-b, \infty] \times [0, \infty]$ for $1 \leq a, b < \infty$.

We have

$$\Big|\int f \mathrm{d}\Psi_t^* - \int f \mathrm{d}\Psi_t\Big| \leq \sum_{n=1}^{M(t)} \Big| f\Big(\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}, \tfrac{F_n - g\left(\log(n\sqrt{\sigma_t})\right)}{g'\left(\log(n\sqrt{\sigma_t})\right)}, \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t)\Big)$$

$$- f\Big(\frac{\frac{1}{\lambda}\log n - \sigma_t}{\sqrt{\sigma_t}}, \tfrac{F_n - g\left(\log(n\sqrt{\sigma_t})\right)}{g'\left(\log(n\sqrt{\sigma_t})\right)}, \xi_n\Big)\Big|$$

$$\leq c_L \sum_{n \in \hat{I}(t)} \Big(\Big|\frac{\tau_n - \tau_n^*}{\sqrt{\sigma_t}}\Big| + \Big|\mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \xi_n\Big|\Big), \tag{2.4}$$

where c_L is the Lipschitz constant of the function f, $\xi_n := \lim_{t\to\infty} \mathrm{e}^{-\gamma t} Y_n(t)$, are i.i.d. copies of ξ, defined in Assumption (A.3), $\tau_n^* = \frac{1}{\lambda}\log n$, and $\hat{I}(t)$ is the random set of indices $n \in \mathbb{N}$ such that

(a) $|\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}| \leq a$ and $\frac{F_n - g(\log(n\sqrt{\sigma_t}))}{g'(\log(n\sqrt{\sigma_t}))} \geq -b$ or

(b) $|\frac{\tau_n^* - \sigma_t}{\sqrt{\sigma_t}}| \leq a$ and $\frac{F_n - g(\log(n\sqrt{\sigma_t}))}{g'(\log(n\sqrt{\sigma_t}))} \geq -b$.

Assume t is large, so that $\sigma_t \leq \frac{t}{3}$ and $\sqrt{\sigma_t} \leq \sigma_t$. For $\varepsilon \in (0, 1/2)$ we denote by $\Upsilon_\varepsilon(t)$ the event that

$$|\tau_n - \tau_n^*| \leq \varepsilon\sqrt{\sigma_t} \quad \text{for all } n \in \mathbb{N}.$$

Assumption (A.1) together with Lemma 7 implies that $\mathbb{P}(\Upsilon_\varepsilon(t)) \to 1$, as $t \to \infty$ for all $\varepsilon > 0$. Now let

$$\bar{I}(t) := \Big\{n \in \mathbb{N} : \tfrac{|\tau_n^* - \sigma_t|}{\sqrt{\sigma_t}} \leq 2a, \text{ and } \tfrac{F_n - g(\log(n\sqrt{\sigma_t}))}{g'(\log(n\sqrt{\sigma_t}))} \geq -b\Big\}.$$

We have that $\hat{I}(t) \subset \bar{I}(t)$ on $\Upsilon_\varepsilon(t)$. Indeed, if (a) and $\Upsilon_\varepsilon(t)$ hold then

$$\frac{|\tau_n^* - \sigma_t|}{\sqrt{\sigma_t}} \leq \frac{|\tau_n^* - \tau_n|}{\sqrt{\sigma_t}} + \frac{|\tau_n - \sigma_t|}{\sqrt{\sigma_t}} \leq \varepsilon + a \leq 2a,$$

and similarly if (b) holds. We now consider the sum on the right-hand side of Equation (2.4), but taken over all $n \in \bar{I}(t)$. First note that, for $n \in \bar{I}(t)$ on $\Upsilon_\varepsilon(t)$, we have

$$\tau_n \leq 2a\sqrt{\sigma_t} + \sigma_t \leq 2a\sigma_t + \sigma_t = \sigma_t(2a+1) \leq \tfrac{t}{2}, \tag{2.5}$$

for t large enough. Since $(\log g)'(\log(n\sqrt{\sigma_t})) \to 0$ as $t \to \infty$, and $g(\log(n\sqrt{\sigma_t})) \to 1$, we have

$$\begin{aligned} F_n &\geq g(\log(n\sqrt{\sigma_t})) - bg'(\log(n\sqrt{\sigma_t})) \qquad (2.6)\\ &= g(\log(n\sqrt{\sigma_t}))\Big(1 - b(\log g)'(\log(n\sqrt{\sigma_t})\Big) \to 1, \end{aligned}$$

as $t \to \infty$.

Recall $\Delta_n(t)$ from Assumption (A.2), $\xi_n = \lim_{t\to\infty} \mathrm{e}^{-\gamma t} Y_n(t)$, and define

$$R_n(t) := \sup_{w>t} \big|\mathrm{e}^{-\gamma w} Y_n(w) - \xi_n\big|.$$

By Assumption (A.3) we have $R_n(t) \to 0$ in probability and, for all t large enough, we have

$$\begin{aligned}
\Big|\mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \xi_n\Big| &\leq \Big|\mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \mathrm{e}^{-\gamma F_n(t-\tau_n)} Y_n(F_n(t-\tau_n))\Big| \\
&\qquad + \Big|\mathrm{e}^{-\gamma F_n(t-\tau_n)} Y_n(F_n(t-\tau_n)) - \xi_n\Big| \\
&\leq \Delta_n(F_n(t-\tau_n)) + R_n(F_n(t-\tau_n)) \\
&\leq \Delta_n\big(\tfrac{t}{2}\big) + R_n\big(\tfrac{t}{2}\big),
\end{aligned}$$

where we have used Equations (2.5) and (2.6). Hence we get that, for sufficiently large t, on $\Upsilon_\varepsilon(t)$,

$$\begin{aligned}
\Big|\int f \mathrm{d}\Psi_t - \int f \mathrm{d}\Psi_t^*\Big| &\leq c_L \sum_{n\in\bar{I}(t)} \Big(\frac{|\tau_n - \tau_n^*|}{\sqrt{\sigma_t}} + \Big|\mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \xi_n\Big|\Big) \\
&\leq c_L \sum_{n\in\bar{I}(t)} \Big(\frac{\sup_n |\tau_n - \tau_n^*|}{\sqrt{\sigma_t}} + \Delta_n\big(\tfrac{t}{2}\big) + R_n\big(\tfrac{t}{2}\big)\Big) \\
&\leq c_L \frac{|\bar{I}(t)| \sup_n |\tau_n - \tau_n^*|}{\sqrt{\sigma_t}} + c_L \sum_{n\in\bar{I}(t)} \Delta_n\big(\tfrac{t}{2}\big) + c_L \sum_{n\in\bar{I}(t)} R_n\big(\tfrac{t}{2}\big).
\end{aligned}$$

By assumption, the random processes $(R_n)_{n\geq1}$ are independent of $(F_n)_{n\geq1}$ and thus also of the random set $\bar{I}(t)$. Recall that, by Lemma 10, $|\bar{I}(t)|$ converges in distribution to a Poisson distribution and hence

$$\lim_{t\to\infty} \sum_{n\in\bar{I}(t)} R_n\big(\tfrac{t}{2}\big) = 0, \quad \text{in probability.}$$

To prove that $\sum_{n\in\bar{I}(t)} \Delta_n\big(\frac{t}{2}\big) \to 0$ in probability as $t\to\infty$ we use Assumption (A.2). Denoting by $\mathbf{F} = (F_m)_{m\in\mathbb{N}}$, we have

$$\begin{aligned}
\mathbb{P}\Big(\sum_{n\in\bar{I}(t)} \Delta_n\big(\tfrac{t}{2}\big) \geq \varepsilon\Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\sum_{n\in\bar{I}(t)} \Delta_n\big(\tfrac{t}{2}\big) \geq \varepsilon \Big| \mathbf{F}\Big)\Big] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}\Big[\mathbb{P}\Big(\exists n \in \bar{I}(t) : \Delta_n\big(\tfrac{t}{2}\big) \geq \frac{\varepsilon}{k} \Big| \mathbf{F}\Big) \mathbf{1}_{\{|\bar{I}(t)|=k\}}\Big] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}\Big[\sum_{n\in\bar{I}(t)} \mathbb{P}\Big(\Delta_n\big(\tfrac{t}{2}\big) \geq \frac{\varepsilon}{k} \Big| \mathbf{F}\Big) \mathbf{1}_{\{|\bar{I}(t)|=k\}}\Big] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}\Big[k \max_{n\in\bar{I}(t)} \mathbb{P}\Big(\Delta_n\big(\tfrac{t}{2}\big) \geq \frac{\varepsilon}{k} \Big| \mathbf{F}\Big) \mathbf{1}_{\{|\bar{I}(t)|=k\}}\Big].
\end{aligned}$$

Now, given $\delta > 0$ pick $K \in \mathbb{N}$ such that, for sufficiently large t,

$$\sum_{k=K+1}^{\infty} \mathbb{E}\Big[k \max_{n \in \bar{I}(t)} \mathbb{P}\Big(\Delta_n\big(\tfrac{t}{2}\big) \geq \frac{\varepsilon}{k}\Big| \mathbf{F}\Big)\mathbf{1}_{\{|\bar{I}(t)|=k\}}\Big] \leq \mathbb{E}\Big[|\bar{I}(t)|\mathbf{1}_{\{|\bar{I}(t)|>K\}}\Big] \leq \frac{\delta}{2},$$

where we use that $|\bar{I}(t)|$ converges in distribution to a Poisson random variable and $\mathbb{E}|\bar{I}(t)|$ converges to its parameter. Since, by definition, $\bar{I}(t) \subseteq I_\kappa(t)$ for $\kappa = 2a$, we get

$$\sum_{k=0}^{K} \mathbb{E}\Big[k \max_{n \in \bar{I}(t)} \mathbb{P}\Big(\Delta_n\big(\tfrac{t}{2}\big) \geq \frac{\varepsilon}{k}\Big| \mathbf{F}\Big)\mathbf{1}_{\{|\bar{I}(t)|=k\}}\Big] \leq K(K+1)\mathbb{E}\Big[\max_{n \in I_\kappa(t)} \mathbb{P}\Big(\Delta_n\big(\tfrac{t}{2}\big) \geq \frac{\varepsilon}{K}\Big| \mathbf{F}\Big)\Big],$$

which converges to zero by (A.2) and dominated convergence.

This shows that $\sum_{n \in \bar{I}(t)} \Delta_n\big(\frac{t}{2}\big) \to 0$ in probability. Summarising, we get

$$\Big|\int f \mathrm{d}\Psi_t - \int f \mathrm{d}\Psi_t^*\Big| \leq c_L |\bar{I}(t)| \frac{\sup_n |\tau_n - \tau_n^*|}{\sqrt{\sigma_t}} + o(1),$$

which converges to zero in probability, as $t \uparrow \infty$. □

Proof of Proposition 9. Let $f\colon (-\infty,\infty) \times (-\infty,\infty] \times [0,\infty] \to \mathbb{R}$ be Lipschitz continuous and compactly supported. Combining Lemmas 10 and 11, together with Slutsky's theorem (see for example [42, Chapter 7.2]) we get the desired result,

$$\int f \mathrm{d}\Psi_t \Rightarrow \int f \mathrm{d}\mathrm{PPP}(\zeta^*) \quad \text{as } t \to \infty,$$

where $\mathrm{PPP}(\zeta^*)$ denotes the Poisson point process with intensity ζ^*. □

2.1.3 Proof of the local convergence result

We are now able to show convergence of the point processes Γ_t.

Proposition 12. *The point process*

$$\Gamma_t = \sum_{n=1}^{M(t)} \delta\Big(\frac{\tau_n - \sigma_t}{\sqrt{\sigma_t}}, \frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}, \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T} Z_n(t)\Big)$$

converges vaguely in distribution on $(-\infty,\infty) \times (-\infty,\infty) \times [0,\infty]$ *to the Poisson point process with intensity*

$$\zeta(\mathrm{d}s,\, \mathrm{d}f,\, \mathrm{d}z) = \lambda \mathrm{e}^{-f} \mathrm{e}^{s^2 a_2 - f a_3} \nu(z \mathrm{e}^{s^2 a_2 - f a_3})\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z.$$

Proof of Proposition 12. Consider the continuous function

$$\phi\colon (s,f,z) \to (s, f, \mathrm{e}^{-s^2 a_2 + f a_3} z),$$

so that $\zeta\circ\phi^{-1}=\zeta^*$. We argue that $\Psi_t\circ\phi^{-1}$ is asymptotically equivalent to Γ_t, that is, for all Lipschitz continuous, compactly supported functions $f\colon(-\infty,\infty)\times(-\infty,\infty)\times[0,\infty]\to\mathbb{R}$,

$$\Big|\int f\mathrm{d}\Psi_t\circ\phi^{-1}-\int f\mathrm{d}\Gamma_t\Big|\to 0\quad\text{in probability, as } t\uparrow\infty.$$

To prove this let f be a Lipschitz continuous function with Lipschitz constant c_L, supported on $K=[-a,a]\times[-b,b]\times[0,\infty]$ for $1\le a,b<\infty$ and abbreviate

$$s_n=\frac{\tau_n-\sigma_t}{\sqrt{\sigma_t}}\qquad\text{and}\qquad f_n=\frac{F_n-g(\log(n\sqrt{\sigma_t}))}{g'(\log(n\sqrt{\sigma_t}))},\qquad\text{for } n\ge 1.$$

We have

$$\begin{aligned}
&\Big|\int f\mathrm{d}\Psi_t\circ\phi^{-1}-\int f\mathrm{d}\Gamma_t\Big|\\
&\quad\le\sum_{n=1}^{M(t)}\Big|f\Big(\frac{\tau_n-\sigma_t}{\sqrt{\sigma_t}},\frac{F_n-g\big(\log(n\sqrt{\sigma_t})\big)}{g\big(\log(n\sqrt{\sigma_t})\big)},\mathrm{e}^{-a_2s_n^2+a_3f_n}\mathrm{e}^{-\gamma F_n(t-\tau_n)}Z_n(t)\Big)\\
&\qquad\qquad-f\Big(\frac{\tau_n-\sigma_t}{\sqrt{\sigma_t}},\frac{F_n-g\big(\log(n\sqrt{\sigma_t})\big)}{g\big(\log(n\sqrt{\sigma_t})\big)},\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1g(\lambda\sigma_t)\log\sigma_t+\gamma T}Z_n(t)\Big)\Big|\\
&\quad\le c_L\sum_{n\in\tilde I(t)}\Big|\mathrm{e}^{-\gamma F_n(t-\tau_n)-a_2s_n^2+a_3f_n}Z_n(t)-\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1g(\lambda\sigma_t)\log\sigma_t+\gamma T}Z_n(t)\Big|,
\end{aligned}\tag{2.7}$$

where $\tilde I(t)$ is the random set of indices $n\in\mathbb{N}$ such that $|s_n|\le a$ and $|f_n|\le b$. We now show that the exponents in (2.7) are asymptotically equivalent, namely

$$-\gamma F_n(t-\tau_n)-a_2s_n^2+a_3f_n=-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1g(\lambda\sigma_t)\log\sigma_t+\gamma T+o(1),\tag{2.8}$$

where the $o(1)$-term does not depend on n. Indeed, combing the definition of s_n and Assumption (A.1), we get

$$\log n=\lambda(\sigma_t+s_n\sqrt{\sigma_t}-T_n),\qquad\text{for } n\ge 1,$$

where we set $T_n=T+\varepsilon_n$. Therefore, we have

$$\begin{aligned}
F_n&=g\big(\log(n\sqrt{\sigma_t})\big)+f_ng'\big(\log(n\sqrt{\sigma_t})\big)\\
&=g\Big(\lambda(\sigma_t+s_n\sqrt{\sigma_t}-T_n)+\tfrac12\log\sigma_t\Big)+f_ng'\Big(\lambda(\sigma_t+s_n\sqrt{\sigma_t}-T_n)+\tfrac12\log\sigma_t\Big).
\end{aligned}$$

Let $x_n:=\lambda s_n\sqrt{\sigma_t}+\frac12\log\sigma_t-\lambda T_n$, so that

$$F_n(t-\tau_n)=\Big(g(\lambda\sigma_t+x_n)+f_ng'(\lambda\sigma_t+x_n)\Big)(t-\sigma_t-s_n\sqrt{\sigma_t}).$$

By the mean value theorem, there exist $c_1, c_2 \in [\lambda\sigma_t, \lambda\sigma_t + x_n]$, such that

$$g(\lambda\sigma_t + x_n) = g(\lambda\sigma_t) + x_n g'(\lambda\sigma_t) + \frac{1}{2}x_n^2 g''(c_1), \quad \text{and} \tag{2.9}$$

$$g'(\lambda\sigma_t + x_n) = g'(\lambda\sigma_t) + x_n g''(c_2). \tag{2.10}$$

Hence, for $n \in \tilde{I}(t)$ we can rewrite

$$\begin{aligned}
&F_n(t-\tau_n) \\
&\quad = \Big(g(\lambda\sigma_t) + x_n g'(\lambda\sigma_t) + \frac{1}{2}x_n^2 g''(c_1) + f_n g'(\lambda\sigma_t) + x_n f_n g''(c_2)\Big)(t - \sigma_t - s_n\sqrt{\sigma_t}) \\
&\quad = g(\lambda\sigma_t)(t-\sigma_t) - g(\lambda\sigma_t)s_n\sqrt{\sigma_t} + \lambda s_n\sqrt{\sigma_t}g'(\lambda\sigma_t)(t-\sigma_t) - \lambda s_n^2\sigma_t g'(\lambda\sigma_t) \\
&\qquad + \Big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\Big)g'(\lambda\sigma_t)(t-\sigma_t - s_n\sqrt{\sigma_t}) + \tfrac{1}{2}x_n^2 g''(c_1)(t-\sigma_t - s_n\sqrt{\sigma_t}) \\
&\qquad + f_n g'(\lambda\sigma_t)(t-\sigma_t) - f_n g'(\lambda\sigma_t)s_n\sqrt{\sigma_t} + f_n x_n g''(c_2)(t-\sigma_t-s_n\sqrt{\sigma_t}).
\end{aligned}$$

Recall that by definition $g'(\lambda\sigma_t)(t-\sigma_t) = \frac{g(\lambda\sigma_t)}{\lambda}$, and $g(\lambda\sigma_t) = 1 + o(1)$ when $t \to \infty$. We get

$$f_n g'(\lambda\sigma_t)(t-\sigma_t) = \frac{f_n}{\lambda} + o(1).$$

By definition $g(\lambda\sigma_t) \uparrow 1$ as $t \uparrow \infty$ and by Lemma 7, we have $\sigma_t = o(t)$ and $g'(\lambda\sigma_t) \sim \frac{1}{\lambda t}$ (see Equations (1.6) and (2.1)). Furthermore, for $n \in \tilde{I}(t)$, Assumption (A.1) implies $T_n = T + \varepsilon_n \to T$, as $t \to \infty$. Combining these with the fact that for all $n \in \tilde{I}(t)$, $|s_n| \le a$ and $|f_n| \le b$, we can show that for all $n \in \tilde{I}(t)$ as $t \to \infty$, the following terms go to zero:

$$\begin{aligned}
&\big|\lambda s_n^2 \sigma_t g'(\lambda\sigma_t)\big| \le \frac{a^2\sigma_t}{t-\sigma_t} = \mathcal{O}\Big(\tfrac{\sigma_t}{t}\Big) = o(1), \\
&\Big|\Big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\Big)g'(\lambda\sigma_t)s_n\sqrt{\sigma_t}\Big| \le \Big|\tfrac{1}{2}\log\sigma_t - \lambda T_n\Big|\frac{a\sqrt{\sigma_t}}{\lambda(t-\sigma_t)} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \sim \Big(\tfrac{1}{2}\log\sigma_t - \lambda T\Big)\frac{a\sqrt{\sigma_t}}{\lambda(t-\sigma_t)} = o(1), \\
&\Big|f_n g'(\lambda\sigma_t)s_n\sqrt{\sigma_t}\Big| \le \frac{ab\sqrt{\sigma_t}}{\lambda(t-\sigma_t)} = \mathcal{O}\Big(\tfrac{\sqrt{\sigma_t}}{t}\Big) = o(1).
\end{aligned}$$

Therefore we can simplify the expression to

$$\begin{aligned}
&F_n(t-\tau_n) \\
&\quad = g(\lambda\sigma_t)(t-\sigma_t) + \frac{g(\lambda\sigma_t)}{2\lambda}\log\sigma_t - g(\lambda\sigma_t)T_n + \frac{f_n}{\lambda} + \frac{1}{2}x_n^2 g''(c_1)(t-\sigma_t-s_n\sqrt{\sigma_t}) \\
&\qquad + f_n x_n g''(c_2)(t-\sigma_t-s_n\sqrt{\sigma_t}) + o(1).
\end{aligned} \tag{2.11}$$

We can write $g(\lambda\sigma_t) = 1 + o(1)$, when $t \to \infty$ and by Assumption (A.1), $T_n = T + o(1)$ in probability, where the $o(1)$-term is with respect to $t \to \infty$ and does not depend on

$n \in \tilde{I}(t)$. Therefore we get

$$g(\lambda\sigma_n)\, T_n = T + o(1) \qquad \text{as } t \to \infty. \tag{2.12}$$

To simplify the last two terms in Equation (2.11), we recall that Lemma 8 implies $g''(c_i) \sim \frac{-\varkappa}{\lambda^2\sigma_t t}$ for $i = 1, 2$. Combing this with the fact that $\sigma_t \to \infty$ as $t \to \infty$ (by Lemma 7), we get for $n \in \tilde{I}(t)$,

$$\begin{aligned}
\left|f_n x_n g''(c_2)(t - \sigma_t - s_n\sqrt{\sigma_t})\right| &= \left|f_n(\lambda s_n\sqrt{\sigma_t} + \tfrac{1}{2}\log\sigma_t - \lambda T_n)g''(c_2)(t - \sigma_t - s_n\sqrt{\sigma_t})\right| \\
&\le \left|b(\lambda a\sqrt{\sigma_t} + \tfrac{1}{2}\log\sigma_t - \lambda T_n)g''(c_2)(t - \sigma_t + a\sqrt{\sigma_t})\right| \\
&= \left|b(\lambda a\sqrt{\sigma_t} + \tfrac{1}{2}\log\sigma_t - \lambda T + o(1))\tfrac{\varkappa(t-\sigma_t+a\sqrt{\sigma_t})}{\lambda^2\sigma_t t} + o(1)\right| \\
&= \mathcal{O}\Big(\tfrac{1}{\sqrt{\sigma_t}}\Big) = o(1), \quad \text{as } t \to \infty.
\end{aligned} \tag{2.13}$$

Consider the penultimate term of Equation (2.11). By the definition of x_n we can rewrite it as follows,

$$\begin{aligned}
&\tfrac{1}{2}x_n^2 g''(c_1)(t - \sigma_t - s_n\sqrt{\sigma_t}) \\
&\quad = \tfrac{1}{2}\big(\lambda s_n\sqrt{\sigma_t} + \tfrac{1}{2}\log\sigma_t - \lambda T_n\big)^2 g''(c_1)(t - \sigma_t - s_n\sqrt{\sigma_t}) \\
&\quad = \tfrac{1}{2}\lambda^2 s_n^2\sigma_t g''(c_1)(t - \sigma_t) - \tfrac{1}{2}\lambda^2 s_n^3\sigma_t^{3/2}g''(c_1) \\
&\qquad\quad + \tfrac{1}{2}\Big(2\lambda s_n\sqrt{\sigma_t}\big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\big) + \big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\big)^2\Big)g''(c_1)(t - \sigma_t - s_n\sqrt{\sigma_t}).
\end{aligned}$$

The first summand is the largest term and by Lemma 8 as $t \to \infty$, it simplifies to

$$\frac{1}{2}\lambda^2 s_n^2\sigma_t g''(c_1)(t - \sigma_t) = -\frac{\lambda^2 s_n^2\sigma_t\varkappa(t - \sigma_t)}{2\lambda^2\sigma_t t} + o(1) = -\frac{1}{2}s_n^2\varkappa + o(1).$$

The second and third summands go to zero as $t \to \infty$, which can be shown using the same results as cited above (Lemmas 7, 8 and Assumption (A.1)). Indeed,

$$\begin{aligned}
&\Big| -\tfrac{1}{2}\lambda^2 s_n^3\sigma_t^{3/2}g''(c_1) \\
&\qquad + \tfrac{1}{2}\Big(2\lambda s_n\sqrt{\sigma_t}\big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\big) + \big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\big)^2\Big)g''(c_1)(t - \sigma_t - s_n\sqrt{\sigma_t})\Big| \\
&\le \Big| -\tfrac{1}{2}\lambda^2 a^3\sigma_t^{3/2}g''(c_1) \\
&\qquad + \tfrac{1}{2}\Big(2\lambda a\sqrt{\sigma_t}\big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\big) + \big(\tfrac{1}{2}\log\sigma_t - \lambda T_n\big)^2\Big)g''(c_1)(t - \sigma_t + a\sqrt{\sigma_t})\Big| \\
&= \Big| \tfrac{1}{2}a^3\sqrt{\sigma_t}\tfrac{\varkappa}{t} \\
&\qquad - \tfrac{1}{2}\Big(2\lambda a\sqrt{\sigma_t}\big(\tfrac{1}{2}\log\sigma_t - \lambda T + o(1)\big) + \big(\tfrac{1}{2}\log\sigma_t - \lambda T + o(1)\big)^2\Big)\tfrac{\varkappa(t-\sigma_t+a\sqrt{\sigma_t})}{\lambda^2\sigma_t t} + o(1)\Big| \\
&= \mathcal{O}\Big(\tfrac{t\sqrt{\sigma_t}\log\sigma_t}{\sigma_t t}\Big) = o(1).
\end{aligned}$$

Therefore, for all $n \in \tilde{I}(t)$, we have

$$\tfrac{1}{2}x_n^2 g''(c_1)(t-\sigma_t - s_n\sqrt{\sigma_t}) = -\tfrac{1}{2}s_n^2\varkappa + o(1), \quad \text{as } t \to \infty. \tag{2.14}$$

Combining (2.12), (2.13) and (2.14), Equation (2.11) becomes

$$F_n(t-\tau_n) \quad = \quad g(\lambda\sigma_t)(t-\sigma_t) + \frac{g(\lambda\sigma_t)}{2\lambda}\log\sigma_t - T - \frac{1}{2}s_n^2\varkappa + \frac{1}{\lambda}f_n + o(1),$$

and thus

$$-\gamma F_n(t-\tau_n) \quad = \quad -\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T + a_2 s_n^2 - a_3 f_n + o(1),$$

where $a_1 = \gamma/2\lambda$, $a_2 = \gamma\varkappa/2$ and $a_3 = \gamma/\lambda$. Rearranging we get Equation (2.8). Substituting Equation (2.8) into (2.7) we get

$$\Big|\int f\mathrm{d}\Psi_t\circ\phi^{-1} \quad - \quad \int f\mathrm{d}\Gamma_t\Big| \le c_L \sum_{n\in\tilde{I}(t)} Z_n(t)\mathrm{e}^{-\gamma F_n(t-\tau_n)}\mathrm{e}^{-a_2 s_n^2 + a_3 f_n}\big|1-\mathrm{e}^{o(1)}\big|.$$

Since the $o(1)$-term does not depend on $n \in \tilde{I}(t)$, we can rewrite it as

$$\Big|\int f\mathrm{d}\Psi_t\circ\phi^{-1} \quad - \quad \int f\mathrm{d}\Gamma_t\Big| \le o\Big(\sum_{n\in\tilde{I}(t)} Z_n(t)\mathrm{e}^{-\gamma F_n(t-\tau_n)}\mathrm{e}^{-a_2 s_n^2 + a_3 f_n}\Big).$$

Furthermore, by definitions of $\Psi_t\circ\phi^{-1}$ and $\tilde{I}(t)$,

$$\begin{aligned}\sum_{n\in\tilde{I}(t)} Z_n(t)\mathrm{e}^{-\gamma F_n(t-\tau_n)}\mathrm{e}^{-a_2 s_n^2 + a_3 f_n} \quad &= \quad \int_0^\infty \mathbf{1}_{|s|\le a}\mathbf{1}_{|f|\le b}\, z\,\mathrm{d}\Psi_t\circ\phi^{-1}(s,\ f\ ,z)\\ &\to \quad \int_0^\infty \mathbf{1}_{|s|\le a}\mathbf{1}_{|f|\le b}\, z\,\mathrm{dPPP}\big(\zeta^*\circ\phi^{-1}\big)\\ &= \quad \int_0^\infty \mathbf{1}_{|s|\le a}\mathbf{1}_{|f|\le b}\, z\,\mathrm{dPPP}(\zeta),\end{aligned}$$

as $t \to \infty$, by Proposition 9.

Recalling the definition of ζ, and substituting $w = z\mathrm{e}^{s^2a_2 - fa_3}$ we get

$$\begin{aligned}&\mathbb{E}\Big[\int_0^\infty z\mathbf{1}_{|s|\le a}\mathbf{1}_{f\ge -b}\,\mathrm{dPPP}(\zeta)\Big]\\ &\qquad = \int_{-a}^{a}\int_{-b}^{b}\int_0^\infty \lambda\mathrm{e}^{-f}\mathrm{e}^{s^2a_2 - fa_3} z\nu\big(z\mathrm{e}^{s^2a_2-fa_3}\big)\,\mathrm{d}s\,\mathrm{d}f\,\mathrm{d}z\\ &\qquad = \int_{-a}^{a}\lambda\mathrm{e}^{-s^2a_2}\,\mathrm{d}s\int_{-b}^{b}\mathrm{e}^{(a_3-1)f}\,\mathrm{d}f\int_0^\infty w\nu(w)\,\mathrm{d}w\\ &\qquad = \lambda\sqrt{\frac{\pi}{a_2}}\,\mathrm{erf}\big(a\sqrt{a_2}\big)\frac{1}{a_3-1}\Big(\mathrm{e}^{(a_3-1)b} - \mathrm{e}^{-(a_3-1)b}\Big)\int_0^\infty w\nu(w)\,\mathrm{d}w\\ &\qquad =: C_1,\end{aligned}$$

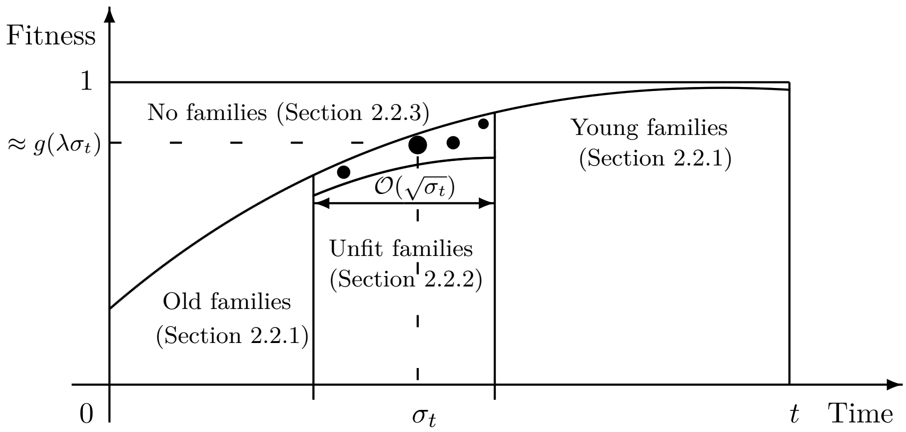
where $\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \mathrm{e}^{-t^2} \mathrm{d}t$. Note that $C_1 < \infty$ since $\int_0^\infty w\nu(w)\,\mathrm{d}w < \infty$ by Assumption (A.3). This implies that

$$\left| \int f \mathrm{d}\Psi_t \circ \phi^{-1} - \int f \mathrm{d}\Gamma_t \right| \leq o(C_1) \to 0, \quad \text{as } t \to \infty,$$

which means that the point process Γ_t is asymptotically equivalent to $\Psi_t \circ \phi^{-1}$. Recall that ζ is the image of ζ^* by the same continuous function, ϕ. By Proposition 9, Ψ_t converges vaguely in distribution on $(-\infty, \infty) \times (-\infty, \infty] \times [0, \infty]$ to $\mathrm{PPP}(\zeta^*)$, so we can conclude that Γ_t converges to $\mathrm{PPP}(\zeta)$ on $(-\infty, \infty) \times (-\infty, \infty) \times [0, \infty]$ as $t \to \infty$. $\square$

2.2 Compactification and completion of the proofs

To deduce Theorem 3 from Proposition 12, one has to control the contribution of the point process near the closed boundaries of $[-\infty, \infty] \times [-\infty, \infty] \times (0, \infty]$. We prove that the families that are born outside of the main window, namely the ones that are unfit or born late, are too small to contribute in the limit. We first consider families which are born either early or late. We then show the negligibility of families lying under the main window by looking at families with small fitness.



2.2.1 Contribution of young and old families

Lemma 13 (Contribution of young and old families)**.** *For every $\eta > 0$ and $\varepsilon > 0$ there exists $v > 1$ such that, for all sufficiently large t, we have*

$$\mathbb{P}\bigg(\max_{n \in \mathcal{I}_t(v)} \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t) \log \sigma_t + \gamma T} Z_n(t) \geq \varepsilon \bigg) \leq \eta,$$

where $\mathcal{I}_t(v) := [0, n_t(-v)] \cup [n_t(v), \infty]$, $n_t(\pm v) := \exp\{\lambda(\sigma_t \pm v\sqrt{\sigma_t})\}$.

Proof. Let $\eta, \varepsilon > 0$. For all $n \geq 1$, we define

$$A_n := \max_{u \geq \tau_n} Z_n(u) \mathrm{e}^{-\gamma F_n(u-\tau_n)}.$$

If there exists $t \geq \tau_n$ such that

$$Z_n(t) \geq \varepsilon \mathrm{e}^{\gamma g(\lambda\sigma_t)(t-\sigma_t)+a_1 g(\lambda\sigma_t)\log\sigma_t - \gamma T}, \tag{2.15}$$

then we get,

$$A_n \geq Z_n(t)\mathrm{e}^{-\gamma F_n(t-\tau_n)} \geq \varepsilon \mathrm{e}^{\gamma g(\lambda\sigma_t)(t-\sigma_t)+a_1 g(\lambda\sigma_t)\log\sigma_t - \gamma T - \gamma F_n(t-\tau_n)}. \tag{2.16}$$

By Assumption (A.1), we have $\tau_n = \frac{1}{\lambda}\log n + T_n$, where $T_n = T + \varepsilon_n$; therefore (2.16) is equivalent to

$$A_n \geq c_{n,t} \mathrm{e}^{-\gamma(1-F_n)T + \gamma F_n \varepsilon_n},$$

where we have set

$$c_{n,t} := \varepsilon \exp\big(\gamma g(\lambda\sigma_t) - \gamma F_n)t + (\gamma F_n - \gamma g(\lambda\sigma_t))\sigma_t + a_1 g(\lambda\sigma_t)\log\sigma_t - \gamma F_n(\sigma_t - \tfrac{1}{\lambda}\log n)\big).$$

Combined with the fact that $\{\max_{n\in\mathbb{N}} Y_n \geq c\} = \bigcup_{n\in\mathbb{N}}\{Y_n \geq c\}$ for any random variables Y_n and $c \in \mathbb{R}$, we get

$$\mathbb{P}\Big(\max_{n\in\mathcal{I}_t(v)} Z_n(t) \geq \varepsilon \mathrm{e}^{\gamma g(\lambda\sigma_t)(t-\sigma_t)+a_1 g(\lambda\sigma_t)\log\sigma_t - \gamma T}\Big) \leq \mathbb{P}\Big(\bigcup_{n\in\mathcal{I}_t(v)} \Big\{A_n \geq c_{n,t}\mathrm{e}^{-\gamma(1-F_n)T+\gamma F_n\varepsilon_n}\Big\}\Big).$$

Moreover, for any $y > 0$, we have

$$\mathbb{P}\Big(\bigcup_{n\in\mathcal{I}_t(v)} \Big\{A_n \geq c_{n,t}\mathrm{e}^{-\gamma(1-F_n)T+\gamma F_n\varepsilon_n}\Big\}\Big) \leq \sum_{n\in\mathcal{I}_t(v)} \mathbb{P}(A_n \geq c_{n,t}\mathrm{e}^{-\gamma y}) + \mathbb{P}(|T| \geq y) + \mathbb{P}\Big(\sup_{n\in\mathcal{I}_t(v)} |\varepsilon_n| \geq y\Big).$$

Since $\varepsilon_n \to 0$ almost surely and $|T|$ is finite, we can fix $y > 0$ large enough, such that $\mathbb{P}(|T| \geq y) \leq \frac{\eta}{3}$ and $\mathbb{P}(\sup_{n\in\mathcal{I}_t(v)} |\varepsilon_n| \geq y) \leq \frac{\eta}{3}$. Consider

$$S := \sum_{n\in\mathcal{I}_t(v)} \mathbb{P}(A_n \geq c_{n,t}\mathrm{e}^{-\gamma y}) = \sum_{n\in\mathcal{I}_t(v)} \mathbb{E}[\mathbb{P}(A_n \geq c_{n,t}\mathrm{e}^{-\gamma y} | (F_m)_{m\in\mathbb{N}})].$$

By Assumption (A.4), $\mathbb{P}(A_n \geq u | (F_m)_{m\in\mathbb{N}}) \leq c_0 \mathrm{e}^{-\eta u}$, so we get

$$
\begin{aligned}
S &\leq c_0 \sum_{n\in\mathcal{I}_t(v)} \mathbb{E}\Big[\exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F_n)t+(\gamma F_n-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\gamma F_n(\sigma_t-\frac{1}{\lambda}\log n)}\Big\}\Big] \\
&\leq c_0 \int_{\mathcal{I}_t(v)} \mathbb{E}\Big[\exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F)t+(\gamma F-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\gamma F(\sigma_t-\frac{1}{\lambda}\log x)}\Big\}\Big]\mathrm{d}x,
\end{aligned}
$$

where F is a random variable of law μ. Let $x = \exp\{\lambda(\sigma_t + w\sqrt{\sigma_t})\}$, therefore we can write

$$
\begin{aligned}
S &\leq c_0 \int_{|w|\geq v} \lambda\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})} \\
&\quad\times \mathbb{E}\Big[\exp\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F)t+(\gamma F-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y+\gamma F w\sqrt{\sigma_t}}\}\Big]\mathrm{d}w \\
&\leq c_0 \int_{|w|\geq v} \lambda\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})} \\
&\quad\times \int_0^1 \mathbb{P}\Big(\exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F)t+(\gamma F-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y+\gamma F w\sqrt{\sigma_t}}\Big\} \geq x\Big)\mathrm{d}x\mathrm{d}w \\
&=: c_0 \int_{|w|\geq v} \lambda\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})} \int_0^1 P(x)\mathrm{d}x\,\mathrm{d}w.
\end{aligned}
$$

Letting $\tilde{x}_0 = 1 + w\sigma_t^{-1/2}$ and substituting into $\mu(x,1) = \exp\{-m(x)\}$, we get

$$
\begin{aligned}
P(x) &= \mathbb{P}\Big(\exp\Big\{-\eta\varepsilon \mathrm{e}^{\gamma g(\lambda\sigma_t)t-\gamma g(\lambda\sigma_t)\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y+\gamma F(-t+\tilde{x}_0\sigma_t)}\Big\} \geq x\Big) \\
&= \mathbb{P}\Big(F \geq (\gamma t-\gamma\tilde{x}_0\sigma_t)^{-1}(\gamma g(\lambda\sigma_t)t-\gamma g(\lambda\sigma_t)\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\log(-\tfrac{1}{\eta\varepsilon}\log x))\Big) \\
&= \mathrm{e}^{-m\Big((\gamma t-\gamma\tilde{x}_0\sigma_t)^{-1}\big(\gamma g(\lambda\sigma_t)t-\gamma g(\lambda\sigma_t)\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\log(-\frac{1}{\eta\varepsilon}\log x)\big)\Big)} \\
&= \mathrm{e}^{-m\Big(\big(1-\tilde{x}_0\frac{\sigma_t}{t}\big)^{-1}\big(g(\lambda\sigma_t)-\frac{g(\lambda\sigma_t)}{t}\sigma_t+\frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\frac{y}{t}-\frac{1}{\gamma t}\log(-\frac{1}{\eta\varepsilon}\log x)\big)\Big)}
\end{aligned}
$$

We can approximate $P(x)$ by

$$
\begin{aligned}
P(x) &= \mathrm{e}^{-m\Big(\big(1+\tilde{x}_0\frac{\sigma_t}{t}\big)\big(g(\lambda\sigma_t)-\frac{\gamma g(\lambda\sigma_t)}{\gamma t}\sigma_t+\frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\frac{y}{t}-\frac{1}{\gamma t}\log(-\frac{1}{\eta\varepsilon}\log x)\big)+\mathcal{O}\big(\frac{\sigma_t}{t}\big)^2\Big)} \\
&= \mathrm{e}^{-m\Big(g(\lambda\sigma_t)+\frac{w\sqrt{\sigma_t}}{t}g(\lambda\sigma_t)+\frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\frac{y}{t}-\frac{1}{\gamma t}\log(-\frac{1}{\eta\varepsilon}\log x)+\mathcal{O}\big(\frac{\sigma_t}{t}\big)^2\Big)}.
\end{aligned}
$$

Lemma 29 (Equation (5.1)) implies

$$
\begin{aligned}
P(x) &= \mathrm{e}^{-m\big(g(\lambda\sigma_t)\big)-m'\big(g(\lambda\sigma_t)\big)\Big(\frac{w\sqrt{\sigma_t}}{t}g(\lambda\sigma_t)+\frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\frac{y}{t}-\frac{1}{\gamma t}\log(-\frac{1}{\eta\varepsilon}\log x)\Big)} \\
&\qquad\times \mathrm{e}^{-\frac{1}{2}m''(c_1)\Big(\frac{w\sqrt{\sigma_t}}{t}g(\lambda\sigma_t)+\frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\frac{y}{t}-\frac{1}{\gamma t}\log(-\frac{1}{\eta\varepsilon}\log x)\Big)^2}.
\end{aligned}
$$

Recall that $m(g(\lambda\sigma_t)) = \lambda\sigma_t$ and $m'(g(\lambda\sigma_t)) = \frac{\lambda(t-\sigma_t)}{g(\lambda\sigma_t)}$. Using Assumption (A5.3), one can show that $m''(g(\lambda\sigma_t)) \sim \frac{\lambda\varkappa t^2}{\sigma_t(g(\lambda\sigma_t))^3}$ as t goes to infinity. Therefore we get

$$\begin{aligned}
P(x) &= \exp\Big\{ -\Big(\lambda\sigma_t + \tfrac{\lambda(t-\sigma_t)}{g(\lambda\sigma_t)}\Big(\tfrac{w\sqrt{\sigma_t}}{t} g(\lambda\sigma_t) + \tfrac{g(\lambda\sigma_t)}{2\lambda t}\log\sigma_t - \tfrac{y}{t} - \tfrac{1}{\gamma t}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big)\Big)\Big\} \\
&\quad \times \exp\Big\{\tfrac{\lambda\varkappa t^2}{2\sigma_t(g(\lambda\sigma_t))^3}\Big(\tfrac{w\sqrt{\sigma_t}}{t} g(\lambda\sigma_t) + \tfrac{g(\lambda\sigma_t)}{2\lambda t}\log\sigma_t - \tfrac{y}{t} - \tfrac{1}{\gamma t}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big)\Big)^2\Big) + o(1)\Big\} \\
&= \sigma_t^{-1/2}\exp\Big\{-\lambda\sigma_t - \lambda w\sqrt{\sigma_t} + \tfrac{\lambda y}{g(\lambda\sigma_t)} + \tfrac{\lambda}{\gamma g(\lambda\sigma_t)}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big) - \tfrac{\lambda\varkappa w^2}{2g(\lambda\sigma_t)} + o(1)\Big\}.
\end{aligned}$$

Hence we get

$$\begin{aligned}
S &\le c_0 \int_{|w|\ge v} \lambda\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t + w\sqrt{\sigma_t})}\sigma_t^{-\frac{1}{2}}\mathrm{e}^{-\lambda(\sigma_t+w\sqrt{\sigma_t}) + \frac{\lambda y}{g(\lambda\sigma_t)} - \frac{\lambda\varkappa w^2}{2g(\lambda\sigma_t)} + o(1)} \\
&\qquad \times \int_0^1 \mathrm{e}^{\frac{\lambda}{\gamma g(\lambda\sigma_t)}\log\left(-\frac{1}{\eta\varepsilon}\log x\right)}\,\mathrm{d}x\,\mathrm{d}w \\
&\le c_0 \int_{|w|\ge v} \lambda \mathrm{e}^{\frac{\lambda y}{g(\lambda\sigma_t)} - \frac{\lambda\varkappa w^2}{2g(\lambda\sigma_t)} + o(1)}\Gamma\Big(\tfrac{\lambda}{\gamma g(\lambda\sigma_t)} + 1\Big)\,\mathrm{d}w \\
&= \mathcal{O}\Big(\int_{|w|\ge v} \exp\Big\{\frac{\lambda y}{g(\lambda\sigma_t)} - \frac{\lambda\varkappa w^2}{2g(\lambda\sigma_t)}\Big\}\,\mathrm{d}w\Big),
\end{aligned}$$

which goes to 0 as v goes to infinity, uniformly for all $t \ge 1$. □

2.2.2 Contribution of unfit families

Lemma 14 (Negligibility of families with small fitnesses)**.** *For every $\eta > 0$ and $\varepsilon > 0$, there exists $\kappa > 0$ such that for all sufficiently large t, we have*

$$\mathbb{P}\Big(\max_{n\le M(t)} \mathbf{1}\Big\{\frac{F_n - g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)} \le -\kappa\Big\}\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T} Z_n(t) \ge \varepsilon\Big) \le \eta.$$

Proof. Let $\varepsilon, \eta > 0$ and $\kappa > 0$. We analyse the event that there exists a family with fitness at most

$$f_n(\kappa) := g\big(\log(n\sqrt{\sigma_t})\big) - \kappa g'\big(\log(n\sqrt{\sigma_t})\big)$$

and size at least $\varepsilon\exp\{\gamma g(\lambda\sigma_t)(t-\sigma_t) + a_1 g(\lambda\sigma_t)\log\sigma_t - \gamma T\}$. Similarly to the proof of Lemma 13, we define, for all $n \ge 1$,

$$A_n := \max_{u\ge\tau_n} Z_n(u)\mathrm{e}^{-\gamma F_n(u-\tau_n)},$$

and as before we define

$$c_{n,t} := \varepsilon\exp\big\{(\gamma g(\lambda\sigma_t) - \gamma F_n)t + (\gamma F_n - \gamma g(\lambda\sigma_t))\sigma_t + a_1 g(\lambda\sigma_t)\log\sigma_t - \gamma F_n(\sigma_t - \tfrac{1}{\lambda}\log n)\big\}.$$

It can be shown that

$$\mathbb{P}\Big(\max_{n\le M(t)} \mathbf{1}_{F_n\le f_n(\kappa)} Z_n(t) \ge \varepsilon \mathrm{e}^{\gamma g(\lambda\sigma_t)(t-\sigma_t)+a_1 g(\lambda\sigma_t)\log\sigma_t}\Big)$$
$$\le \sum_{n=1}^{\infty} \mathbb{P}(A_n \mathbf{1}_{F_n\le f_n(\kappa)} \ge c_{n,t}\mathrm{e}^{-\gamma y}) + \mathbb{P}(|T|\ge y) + \mathbb{P}(\sup_{n\in\mathbb{N}} |\varepsilon_n| \ge y),$$

where $y>0$ is large enough, so that $\mathbb{P}(|T|\ge y)\le \frac{\eta}{3}$ and $\mathbb{P}(\sup_{n\in\mathbb{N}}|\varepsilon_n|\ge y)\le\frac{\eta}{3}$. Set

$$S := \sum_{n=1}^{\infty} \mathbb{P}(A_n \mathbf{1}_{F_n\le f_n(\kappa)} \ge c_{n,t}\mathrm{e}^{-\gamma y}) = \sum_{n=1}^{\infty} \mathbb{E}\big[\mathbf{1}_{F_n\le f_n(\kappa)} \mathbb{P}(A_n \ge c_{n,t}\mathrm{e}^{-\gamma y}\big|(F_m)_{m\in\mathbb{N}})\big].$$

By Assumption (A.4), $\mathbb{P}(A_n\ge u|(F_m)_{m\in\mathbb{N}})\le c_0\mathrm{e}^{-\eta u}$, which implies

$$\begin{aligned}
S \le{}& c_0 \sum_{n=1}^{\infty} \mathbb{E}\Big[\mathbf{1}_{F_n\le f_n(\kappa)} \\
&\qquad \times \exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F_n)t+(\gamma F_n-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\gamma F_n(\sigma_t-\frac{1}{\lambda}\log n)}\Big\}\Big] \\
\le{}& c_0 \int_0^{\infty} \mathbb{E}\Big[\mathbf{1}_{F\le f_x(\kappa)} \\
&\qquad \times \exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F)t+(\gamma F-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\gamma F(\sigma_t-\frac{1}{\lambda}\log x)}\Big\}\Big]\mathrm{d}x,
\end{aligned}$$

where F is a random variable of law μ. Let $x=\exp\{\lambda(\sigma_t+w\sqrt{\sigma_t})\}$ and $\hat{f}_w(\kappa) := f_{\exp\{\lambda(\sigma_t+w\sqrt{\sigma_t})\}}(\kappa)$, therefore

$$\begin{aligned}
S \le{}& c_0 \int_{-\infty}^{\infty} \lambda\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})} \\
&\times \mathbb{E}\Big[\mathbf{1}_{F\le \hat{f}_w(\kappa)} \exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F)t+(\gamma F-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y+\gamma F w\sqrt{\sigma_t}}\Big\}\Big]\mathrm{d}w.
\end{aligned}$$

Denoting by $\tilde{x}_0 := 1+w{\sigma_t}^{-1/2}$ and

$$E := \mathbb{E}\Big[\mathbf{1}_{F\le \hat{f}_w(\kappa)} \exp\Big\{-\eta\varepsilon \mathrm{e}^{(\gamma g(\lambda\sigma_t)-\gamma F)t+(\gamma F-\gamma g(\lambda\sigma_t))\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y+\gamma F w\sqrt{\sigma_t}}\Big\}\Big],$$

we get

$$\begin{aligned}
E ={}& \int_0^1 \mathbb{P}\Big(F\le \hat{f}_w(\kappa);\ \exp\Big\{-\eta\varepsilon \mathrm{e}^{\gamma g(\lambda\sigma_t)t-\gamma g(\lambda\sigma_t)\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y+\gamma F(-t+\tilde{x}_0\sigma_t)}\Big\}\ge x\Big)\mathrm{d}x \\
={}& \int_0^1 \mathbb{P}\Big(\tfrac{\gamma g(\lambda\sigma_t)t-\gamma g(\lambda\sigma_t)\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\log\left(\frac{1}{\eta\varepsilon}\log\left(\frac{1}{x}\right)\right)}{\gamma(t-\tilde{x}_0\sigma_t)} \le F\le \hat{f}_w(\kappa)\Big)\,\mathrm{d}x.
\end{aligned}$$

Note that

$$\begin{aligned}
f_x :={}& \tfrac{\gamma g(\lambda\sigma_t)t-\gamma g(\lambda\sigma_t)\sigma_t+a_1 g(\lambda\sigma_t)\log\sigma_t-\gamma y-\log\left(\frac{1}{\eta\varepsilon}\log\left(\frac{1}{x}\right)\right)}{\gamma(t-\tilde{x}_0\sigma_t)} \\
={}& \Big(g(\lambda\sigma_t)-\tfrac{g(\lambda\sigma_t)}{t}\sigma_t+\tfrac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\tfrac{y}{t}-\tfrac{1}{\gamma t}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big)\Big)\Big(1+\tfrac{\tilde{x}_0}{t}\sigma_t+\mathcal{O}\big(\tfrac{\sigma_t}{t}\big)^2\Big) \\
={}& g(\lambda\sigma_t)-\big(g(\lambda\sigma_t)-g(\lambda\sigma_t)\tilde{x}_0\big)\tfrac{\sigma_t}{t}+\tfrac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t-\tfrac{y}{t}-\tfrac{1}{\gamma t}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big)+o\big(\tfrac{1}{t}\big).
\end{aligned}$$

We have $E = \int_0^1 \big(\mu(f_x,1) - \mu(\hat{f}_w(\kappa),1)\big)\mathrm{d}x$. By Lemma 29 (Equation (5.1)),

$$\begin{aligned}
\mu(f_x,1) &= \mathrm{e}^{-m\Big(g(\lambda\sigma_t) - \big(g(\lambda\sigma_t) - g(\lambda\sigma_t)\tilde{x}_0\big)\frac{\sigma_t}{t} + \frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t - \frac{y}{t} - \frac{1}{\gamma t}\log\big(-\frac{1}{\eta\varepsilon}\log x\big) + o\big(\frac{1}{t}\big)\Big)} \\
&= \mathrm{e}^{-m\Big(g(\lambda\sigma_t) + g(\lambda\sigma_t)\frac{w\sqrt{\sigma_t}}{t} + \frac{g(\lambda\sigma_t)}{2\lambda t}\log\sigma_t - \frac{y}{t} - \frac{1}{\gamma t}\log\big(-\frac{1}{\eta\varepsilon}\log x\big) + o\big(\frac{1}{t}\big)\Big)} \\
&= \mathrm{e}^{-m\big(g(\lambda\sigma_t)\big) - m'\big(g(\lambda\sigma_t)\big)\Big(g(\lambda\sigma_t)\frac{w\sqrt{\sigma_t}}{t} + \frac{g(\lambda\sigma_t)}{2\lambda t}\log\sigma_t - \frac{y}{t} - \frac{1}{\gamma t}\log\big(-\frac{1}{\eta\varepsilon}\log x\big)\Big)} \\
&\qquad\qquad \times \mathrm{e}^{-\frac{1}{2}m''(c_3)\Big(g(\lambda\sigma_t)\frac{w\sqrt{\sigma_t}}{t} + \frac{g(\lambda\sigma_t)}{2\lambda t}\log\sigma_t - \frac{y}{t} - \frac{1}{\gamma t}\log\big(-\frac{1}{\eta\varepsilon}\log x\big)\Big)^2} \\
&= \mathrm{e}^{-\lambda\sigma_t - \lambda w\sqrt{\sigma_t} - \log\sqrt{\sigma_t} + \frac{\lambda y}{g(\lambda\sigma_t)} + \frac{\lambda}{\gamma g(\lambda\sigma_t)}\log\big(-\frac{1}{\eta\varepsilon}\log x\big) - \frac{\lambda\varkappa w^2}{2g(\lambda\sigma_t)} + o(1)},
\end{aligned}$$

since $m''(g(\lambda\sigma_t)) \sim \frac{\lambda\varkappa t^2}{\sigma_t(g(\lambda\sigma_t))^3}$, by Assumption (A5.3). Using Lemma 29 (Equation (5.2)), and the fact that $m'(g(x))g'(x) = 1$ for all $x > 0$, we get

$$\begin{aligned}
&\mu(\hat{f}_w(\kappa),1) \\
&\quad = \exp\Big\{ - m\Big(g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t}) - \kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big)\Big\} \\
&\quad = \exp\Big\{ - m\Big(g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big) + m'\Big(g(\lambda\tilde{x}_0\sqrt{\sigma_t} + \log\sqrt{\sigma_t})\Big)\kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sigma_t)\Big\} \\
&\qquad\qquad\qquad \times \exp\Big\{ - \tfrac{1}{2}m''(c_4)\Big(\kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big)^2\Big\} \\
&\quad = \exp\Big\{ - \lambda\tilde{x}_0\sigma_t - \log\sqrt{\sigma_t} + \kappa - \tfrac{\lambda\varkappa t^2}{2\sigma_t(g(\lambda\sigma_t))^3}\Big(\kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big)^2\Big\} \\
&\quad = \exp\Big\{ - \lambda\tilde{x}_0\sigma_t - \log\sqrt{\sigma_t} + \kappa + o(1)\Big\},
\end{aligned}$$

as $t \to \infty$. This last equality holds in view of Lemma 29 (Equation (5.3)), since

$$\begin{aligned}
&\frac{\lambda\varkappa t^2}{2(g(\lambda\sigma_t))^3\sigma_t}\Big(\kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big)^2 \\
&\qquad = \frac{\lambda\varkappa t^2}{2(g(\lambda\sigma_t))^3\sigma_t}\kappa^2\Big(g'(\lambda\sigma_t) + g''(c_2)(\lambda w\sigma_t + \log\sqrt{\sigma_t})\Big)^2 \\
&\qquad = \frac{\lambda\varkappa t^2}{2(g(\lambda\sigma_t))^3\sigma_t}\kappa^2\Big(\frac{g(\lambda\sigma_t)}{\lambda(t-\sigma_t)} - \frac{\varkappa}{\sigma_t t}(\lambda w\sqrt{\sigma_t} + \log\sqrt{\sigma_t})\Big)^2 \\
&\qquad = \mathcal{O}(\sigma_t^{-1}) = o(1).
\end{aligned}$$

For $E > 0$, we need $\mu(f_x,1) > \mu(\hat{f}_w(\kappa),1)$, which holds if and only if

$$x \le \exp\Big\{ - \frac{\varepsilon}{2}\exp\Big\{\frac{\gamma}{\lambda}g(\lambda\sigma_t)\Big(\kappa + \frac{\lambda\varkappa w^2}{2g(\lambda\sigma_t)} - \frac{\lambda y}{g(\lambda\sigma_t)}\Big)\Big\}\Big\} =: f_1.$$

Since $g(\lambda\sigma_t) \to 1$ as $t \to \infty$, we get

$$f_1 = \exp\Big\{ - \eta\varepsilon\exp\Big\{\frac{\gamma}{\lambda}\Big(\kappa + \frac{\lambda\varkappa w^2}{2}\Big) - \lambda y + o(1)\Big\}\Big\}.$$

Hence we can rewrite E as

$$
\begin{aligned}
E &= (1+o(1))\mathrm{e}^{-\lambda\sigma_t-\lambda w\sqrt{\sigma_t}-\log\sqrt{\sigma_t}}\int_0^{f_1}\left(\mathrm{e}^{\frac{\lambda}{\gamma}\log\left(-\frac{1}{\eta\varepsilon}\log x\right)-\frac{\lambda\varkappa w^2}{2}+\lambda y}-\mathrm{e}^{\kappa}\right)\mathrm{d}x \\
&= (1+o(1))\mathrm{e}^{-\lambda\sigma_t-\lambda w\sqrt{\sigma_t}-\log\sqrt{\sigma_t}}\Big(\mathrm{e}^{\lambda y-\frac{\lambda\varkappa w^2}{2}}\left(\tfrac{1}{\eta\varepsilon}\right)^{\frac{\lambda}{\gamma}}\int_0^{f_1}\left(\log\tfrac{1}{x}\right)^{\frac{\lambda}{\gamma}}\mathrm{d}x-\int_0^{f_1}\mathrm{e}^{\kappa}\,\mathrm{d}x\Big) \\
&= (1+o(1))\mathrm{e}^{-\lambda\sigma_t-\lambda w\sqrt{\sigma_t}-\log\sqrt{\sigma_t}}\Big(\mathrm{e}^{\lambda y-\frac{\lambda\varkappa w^2}{2}}\left(\tfrac{1}{\eta\varepsilon}\right)^{\frac{\lambda}{\gamma}}\Gamma\Big(\frac{\lambda}{\gamma}+1,\eta\varepsilon\mathrm{e}^{\frac{\gamma}{\lambda}(\kappa+\frac{\lambda\varkappa w^2}{2})}\Big) \\
&\qquad\qquad -\exp\left\{\kappa-\eta\varepsilon\mathrm{e}^{\frac{\gamma}{\lambda}(\kappa+\frac{\lambda\varkappa w^2}{2}-\lambda y)}\right\}\Big),
\end{aligned}
$$

where $\Gamma(s,x)=\int_x^\infty z^{s-1}\mathrm{e}^{-z}\mathrm{d}z$ is the upper incomplete gamma function. So we get

$$
\begin{aligned}
&S \le (1+o(1))c_0 \\
&\int_{-\infty}^{\infty}\lambda\Big(\mathrm{e}^{\lambda y-\frac{\lambda\varkappa w^2}{2}}\left(\tfrac{1}{\eta\varepsilon}\right)^{\frac{\lambda}{\gamma}}\Gamma\big(\tfrac{\lambda}{\gamma}+1,\eta\varepsilon\mathrm{e}^{\frac{\gamma}{\lambda}(\kappa+\frac{\lambda\varkappa w^2}{2}-\lambda y)}\big)-\exp\Big\{\kappa-\eta\varepsilon\mathrm{e}^{\frac{\gamma}{\lambda}(\kappa+\frac{\lambda\varkappa w^2}{2}-\lambda y)}\Big\}\Big)\mathrm{d}w.
\end{aligned}
$$

Since $\frac{\Gamma(s,x)}{x^{s-1}\mathrm{e}^{-x}}\to 1$ as $x\to\infty$, as $\kappa\to\infty$ we have

$$
\Gamma\big(\tfrac{\lambda}{\gamma}+1,\eta\varepsilon\mathrm{e}^{\frac{\gamma}{\lambda}(\kappa+\frac{\lambda\varkappa w^2}{2}-\lambda y)}\big)\sim(\eta\varepsilon)^{\frac{\lambda}{\gamma}}\exp\Big\{\kappa+\frac{\lambda\varkappa w^2}{2}-\lambda y-\eta\varepsilon\mathrm{e}^{\frac{\gamma}{\lambda}(\kappa+\frac{\lambda\varkappa w^2}{2}-\lambda y)}\Big\},
$$

and so $S\to 0$. $\square$

2.2.3 Contribution of old and fit families

Lemma 15 (Absence of fit families above the "window")**.** *For every $\varepsilon>0$ and $\nu>0$, there exists $\kappa>0$ such that for all sufficiently large t, we have*

$$
\mathbb{P}\Big(\max_{n\in\mathcal{I}_t^c(v)}\frac{F_n-g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}\le\kappa\Big)\ge 1-\varepsilon, \tag{2.17}
$$

where $\mathcal{I}_t^c(v)=[n_t(-v),n_t(v)]$, $n_t(\pm v):=\exp\{\lambda(\sigma_t\pm v\sqrt{\sigma_t})\}$.

Proof. Let $\varepsilon,v>0$ and $\kappa>0$. We have

$$
\begin{aligned}
\mathbb{P}\Big(\max_{n\in\mathcal{I}_t^c(v)}\Big(\frac{F_n-g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}\Big)\le\kappa\Big) &= \prod_{n_t(-v)}^{n_t(v)}\mathbb{P}\Big(F_n\le g\big(\log(n\sqrt{\sigma_t})\big)+\kappa g'\big(\log(n\sqrt{\sigma_t})\big)\Big) \\
&= \prod_{n_t(-v)}^{n_t(v)}\Big(1-\mu\Big(g\big(\log(n\sqrt{\sigma_t})\big)+\kappa g'\big(\log(n\sqrt{\sigma_t})\big),1\Big)\Big).
\end{aligned}
$$

Using the fact that $\mathrm{e}^{-\mu(x,1)}=1-\mu(x,1)+o(\mu(x,1))$ when $x\to 1$, we get that when $t\to\infty$,

$$
\mathbb{P}\Big(\max_{n\in\mathcal{I}_t^c(v)}\Big(\frac{F_n-g\big(\log(n\sqrt{\sigma_t})\big)}{g'\big(\log(n\sqrt{\sigma_t})\big)}\Big)\le\kappa\Big)\sim\exp\Big\{-\sum_{n_t(-v)}^{n_t(v)}\mu\Big(g\big(\log(n\sqrt{\sigma_t})\big)+\kappa g'\big(\log(n\sqrt{\sigma_t})\big),1\Big)\Big\}.
$$

Recall that $\mu(x,1)=\mathrm{e}^{-m(x)}$, which implies that

$$\mathbb{P}\Big(\max_{n\in\mathcal{I}_t^c(v)}\Big(\tfrac{F_n-g\left(\log(n\sqrt{\sigma_t})\right)}{g'\left(\log(n\sqrt{\sigma_t})\right)}\Big)\leq\kappa\Big)\sim\exp\Big\{-\int_{n_t(-v)}^{n_t(v)}\mathrm{e}^{-m\left(g\left(\log(x\sqrt{\sigma_t})\right)+\kappa g'\left(\log(x\sqrt{\sigma_t})\right)\right)}\mathrm{d}x\Big\}.$$

Using the change of variables with $x=\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})}=n_t(w)$, we get

$$\begin{aligned}\mathbb{P}\Big(&\max_{n\in\mathcal{I}_t^c(v)}\Big(\tfrac{F_n-g\left(\log(n\sqrt{\sigma_t})\right)}{g'\left(\log(n\sqrt{\sigma_t})\right)}\Big)\leq\kappa\Big)\\&\sim\exp\Big\{-\int_{-v}^{v}\mathrm{e}^{-m\left(g\left(\log(n_t(w)\sqrt{\sigma_t})\right)+\kappa g'\left(\log(n_t(w)\sqrt{\sigma_t})\right)\right)}\lambda\sqrt{\sigma_t}n_t(w)\mathrm{d}w\Big\}.\end{aligned}$$

By the same technique as in Lemma 15(a), we get that there exists

$$c_6\in[g\big(\log(n_t(w)\sqrt{\sigma_t})\big),g\big(\log(n_t(w)\sqrt{\sigma_t})\big)+\kappa g'\big(\log(n_t(w)\sqrt{\sigma_t})\big)]$$

such that

$$\begin{aligned}m\Big(&g\big(\log(n_t(w)\sqrt{\sigma_t})\big)+\kappa g'\big(\log(n_t(w)\sqrt{\sigma_t})\big)\Big)\\&=m\Big(g\big(\log(n_t(w)\sqrt{\sigma_t})\big)\Big)+\kappa m'\Big(g\big(\log(n_t(w)\sqrt{\sigma_t})\big)\Big)g'\big(\log(n_t(w)\sqrt{\sigma_t})\big)\Big)\\&\qquad+\tfrac{1}{2}m''(c_6)\Big(\kappa g'\big(\log(n_t(w)\sqrt{\sigma_t})\big)\Big)^2\\&=\log(n_t(w)\sqrt{\sigma_t}))+\kappa+\mathcal{O}(\sigma_t^{-1})=\lambda(\sigma_t+w\sqrt{\sigma_t})+\log(\sqrt{\sigma_t})+\kappa+o(1),\end{aligned}$$

where we have used that $m(g(x))=x$ and hence $m'(g(x))g'(x)=1$ for all $x>0$. We also used the fact that $m''(c_6)(\kappa g'(\log(n_t(w)\sqrt{\sigma_t})))^2\to 0$ as $t\to\infty$, by Assumption (A5.2). Therefore, the integral becomes

$$\begin{aligned}\int_{-v}^{v}&\mathrm{e}^{-m\left(g\left(\lambda(\sigma_t+w\sqrt{\sigma_t})+\kappa g'\left(\lambda(\sigma_t+w\sqrt{\sigma_t})\right)\right)\right)}\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})}\,\mathrm{d}w\\&=\int_{-v}^{v}\mathrm{e}^{-\lambda(\sigma_t+w\sqrt{\sigma_t})-\log(\sqrt{\sigma_t})-\kappa-\mathcal{O}(\sigma_t^{-1})}\sqrt{\sigma_t}\mathrm{e}^{\lambda(\sigma_t+w\sqrt{\sigma_t})}\,\mathrm{d}w\\&=\int_{-v}^{v}\mathrm{e}^{-\kappa-\mathcal{O}(\sigma_t^{-1})}\,\mathrm{d}w\sim 2v\mathrm{e}^{-\kappa}.\end{aligned}$$

Therefore, we get

$$\mathbb{P}\Big(\max_{n\in\mathcal{I}_t^c(v)}\Big(\tfrac{F_n-g\left(\log(n\sqrt{\sigma_t})\right)}{g'\left(\log(n\sqrt{\sigma_t})\right)}\Big)\leq\kappa\Big)\quad\sim\quad\exp\big\{-2v\mathrm{e}^{-\kappa}\big\}\to 1,\quad\text{as }\kappa\to\infty.$$

□

2.2.4 Proof of Theorem 3

Let $\eta, \varepsilon > 0$. By Lemma 14 there exists $\kappa_1 = \kappa_1(\varepsilon, \eta)$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t([-\infty,\infty]\times[-\infty,-\kappa_1]\times(\varepsilon,\infty]) = 0\Big) \geq 1-\eta.$$

By Lemma 13 there exists $v = v(\varepsilon, \eta) > 1$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t([-\infty,-v]\cup[v,\infty]\times[-\infty,\infty]\times(\varepsilon,\infty]) = 0\Big) \geq 1-\eta.$$

By Lemma 15 there exists $\kappa_2 = \kappa_2(\varepsilon, \eta)$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t([-v,v]\times[\kappa_2,\infty]\times(\varepsilon,\infty]) = 0\Big) \geq 1-\eta.$$

Finally, Proposition 12 gives that Γ_t converges on $(-v,v)\times(-\kappa_1,\kappa_2)\times(\varepsilon,\infty]$ to the Poisson process with intensity measure ζ. Combining these four facts and using that $\eta > 0$ is arbitrarily small, we get convergence on $[-\infty,\infty]\times[-\infty,\infty]\times(\varepsilon,\infty]$. As this holds for all $\varepsilon > 0$, the proof is complete.

2.2.5 Proof of Corollary 4

(i) We fix $x > 0$ and $B := [-\infty,\infty]\times[-\infty,\infty]\times[x,\infty]$. By Theorem 3, we get that, as $t \uparrow \infty$,

$$\sum_{n=1}^{M(t)} \mathbf{1}_B\Big(\tfrac{\tau_n-\sigma_t}{\sqrt{\sigma_t}}, \tfrac{F_n-g(\log(n\sqrt{\sigma_t}))}{g'(\log(n\sqrt{\sigma_t}))}, \mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} Z_n(t)\Big) \Rightarrow \mathrm{Poisson}\Big(\int_B \mathrm{d}\zeta\Big),$$

since B is a compact set. Hence, as $t \uparrow \infty$,

$$\begin{aligned}
&\mathbb{P}\Big(\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} \max_{n\in\{1,\dots,M(t)\}} Z_n(t) \geq x\Big) \\
&\quad\to \mathbb{P}\Big(\mathrm{Poisson}\Big(\int_B \mathrm{d}\zeta\Big) \geq 1\Big) = 1 - \mathbb{P}\Big(\mathrm{Poisson}\Big(\int_B \mathrm{d}\zeta\Big) = 0\Big) = 1 - \exp\Big(-\int_B \mathrm{d}\zeta\Big).
\end{aligned} \tag{2.18}$$

Note that

$$\begin{aligned}
\int_B \mathrm{d}\zeta &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_x^{\infty} \lambda \mathrm{e}^{-f}\mathrm{e}^{s^2a_2-fa_3}\nu(z\mathrm{e}^{s^2a_2-fa_3})\,\mathrm{d}z\,\mathrm{d}f\,\mathrm{d}s \\
&= \lambda\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{x\mathrm{e}^{s^2a_2-fa_3}}^{\infty} \mathrm{e}^{-f}\nu(w)\,\mathrm{d}w\,\mathrm{d}f\,\mathrm{d}s \\
&= \lambda\int_{-\infty}^{\infty}\int_0^{\infty}\int_{\frac{1}{a_3}(s^2a_2-\log\frac{w}{x})}^{\infty} \mathrm{e}^{-f}\nu(w)\,\mathrm{d}f\,\mathrm{d}w\,\mathrm{d}s \\
&= \lambda\Big(\int_{-\infty}^{\infty} \mathrm{e}^{-\frac{a_2}{a_3}s^2}\mathrm{d}s\Big)\Big(\int_0^{\infty}\nu(w)(\tfrac{w}{x})^{\frac{1}{a_3}}\,\mathrm{d}w\Big) \\
&= \lambda\sqrt{\pi\frac{a_3}{a_2}}\Big(\int_0^{\infty}\nu(w)w^{\frac{1}{a_3}}\,\mathrm{d}w\Big)x^{-\frac{1}{a_3}}.
\end{aligned} \tag{2.19}$$

Recall that $a_2 = \gamma\varkappa/2$ and $a_3 = \frac{\gamma}{\lambda}$. Thus the right hand side in (2.18) is $1-\exp(-s^\eta x^{-\eta})$, for

$$s^\eta = \sqrt{\tfrac{2\pi\lambda}{\varkappa}} \int_0^\infty \nu(w) w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w, \quad \text{and } \eta = \frac{\lambda}{\gamma}.$$

In summary, for all $x > 0$, we have

$$\mathbb{P}\Big(\mathrm{e}^{-\gamma g(\lambda\sigma_t)(t-\sigma_t)-a_1 g(\lambda\sigma_t)\log\sigma_t+\gamma T} \max_{n\in\{1,\dots,M(t)\}} Z_n(t) \le x\Big) \quad \to \quad \mathrm{e}^{-\left(\frac{x}{s}\right)^{-\frac{\lambda}{\gamma}}} = \mathbb{P}(W \le x),$$

where $W \sim$ Fréchet $\left(\frac{\lambda}{\gamma}, s\right)$.

(ii) By Theorem 3 the random variable $\frac{S(t)-\sigma_t}{\sqrt{\sigma_t}}$ converges to a random variable U with density

$$\int_{-\infty}^\infty \int_0^\infty \mathrm{e}^{-\zeta([\infty,+\infty]\times[-\infty,+\infty]\times[z,+\infty])}\zeta(s,\mathrm{d}f,\mathrm{d}z).$$

We recall from above that

$$\zeta([-\infty,+\infty]\times[-\infty,+\infty]\times[z,+\infty]) = \lambda\sqrt{\pi\frac{a_3}{a_2}}\Big(\int_0^\infty \nu(w)w^{\frac{1}{a_3}}\,\mathrm{d}w\Big)z^{-\frac{1}{a_3}} =: c_6 z^{-\frac{1}{a_3}}.$$

We get, substituting $u = z\mathrm{e}^{s^2 a_2 - f a_3}$,

$$\begin{aligned}&\int_{-\infty}^\infty \int_0^\infty \mathrm{e}^{-\zeta([-\infty,+\infty]\times[-\infty,+\infty]\times[z,+\infty])}\,\mathrm{d}\zeta(s,f,z)\\ &\qquad = \lambda\int_0^\infty \nu(u)\int_{-\infty}^\infty \exp\Big\{-f - c_6 u^{-\frac{1}{a_3}}\mathrm{e}^{s^2\frac{a_2}{a_3}}\mathrm{e}^{-f}\Big\}\,\mathrm{d}f\,\mathrm{d}u\,\mathrm{d}s.\end{aligned}$$

Integrating with respect to f and simplifying, gives us

$$\begin{aligned}\int_{-\infty}^\infty \int_0^\infty \mathrm{e}^{-\zeta([-\infty,+\infty]\times[-\infty,+\infty]\times[z,+\infty])}\mathrm{d}\zeta(s,f,z) &= \frac{\lambda}{c_6}\mathrm{e}^{-s^2\frac{a_2}{a_3}}\,\mathrm{d}s\int_0^\infty \nu(u)u^{\frac{1}{a_3}}\,\mathrm{d}u\\ &= \frac{1}{\sqrt{\frac{2\pi}{\lambda\varkappa}}}\mathrm{e}^{-s^2\frac{\lambda\varkappa}{2}}\,\mathrm{d}s.\end{aligned}$$

Chapter 3

μ in MDA(Weibull)

In this chapter we prove Theorem 1 and Corollary 2. The proofs are based on similar ideas to the ones in [26] and in Chapter 2, but they are more general than the former and less involved than the latter, due to the nice properties of distributions lying in the MDA(Weibull).

3.1 Local Convergence

3.1.1 Approximation by a simpler point process

In this section we prove the following proposition.

Proposition 16. *The point process*

$$\hat{\Gamma}_t = \sum_{n=1}^{M(t)} \delta\big(\tau_n - \sigma_t, (t-\tau_n)(1-F_n), \mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t)\big),$$

converges vaguely in distribution on $(-\infty, \infty) \times [0, \infty) \times [0, \infty]$ *to the Poisson point process with intensity*

$$\mathrm{d}\zeta(s, f, z) = \alpha f^{\alpha-1} \lambda \mathrm{e}^{\lambda s} \mathrm{e}^{\gamma(s+f)} \nu\big(z\, \mathrm{e}^{\gamma(s+f)}\big)\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z.$$

Proposition 16 follows from the result below.

Proposition 17. *We have vague convergence in distribution of the point process*

$$\Psi_t = \sum_{n=1}^{M(t)} \delta\big(\tau_n - \sigma_t, (t-\tau_n)(1-F_n), \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t)\big),$$

to the Poisson point process with intensity

$$\mathrm{d}\zeta^*(s, f, z) = \alpha f^{\alpha-1} \lambda \mathrm{e}^{\lambda s} \nu(z)\, \mathrm{d}s\, \mathrm{d}f\, \mathrm{d}z,$$

on $(-\infty, \infty) \times [0, \infty) \times [0, \infty]$.

Proof of Proposition 16 given Proposition 17 holds. Convergence in Proposition 16 follows from the fact that the point process $\hat{\Gamma}_t$ is the image of Ψ_t by the continuous function $\phi\colon (s, f, z) \to (s, f, \mathrm{e}^{-\gamma(s+f)}z)$, and that ζ is the image of ζ^* by the same continuous function. The exponent of the third coordinate of Ψ_t under ϕ becomes

$$-\gamma F_n(t-\tau_n) - \gamma(s+f) = -\gamma F_n(t-\tau_n) - \gamma(\tau_n - \sigma_t + (t-\tau_n)(1-F_n)) = -\gamma(t-\sigma_t),$$

as required. □

We now prove Proposition 17, by approximating Ψ_t with a point process

$$\Psi_t^* = \sum_{n \in \mathbb{N}} \delta\big(\tau_n - \sigma_t, t(1-F_n), \xi_n\big),$$

where the rescaled family sizes $\mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t)$ are replaced by $\xi_n := \lim_{t \to \infty} \mathrm{e}^{-\gamma t} Y_n(t)$, which are i.i.d. copies of ξ defined in Assumption (A.3).

In our choice of the approximating process Ψ_t^* the components are decoupled, so it is easier to study. We then prove Proposition 17 in two steps, namely we prove that

(1) the approximation process Ψ_t^* converges vaguely to the Poisson point process of intensity ζ^* (see Lemma 18), and

(2) Ψ_t^* is close enough to Ψ_t to imply Proposition 17 (see Lemma 19).

Lemma 18. *Under Assumption (B.5)* $(\Psi_t^*)_{t \geq 0}$ *converges vaguely in distribution on* $[-\infty, \infty) \times [0, \infty) \times [0, \infty]$ *to the Poisson point process with intensity* ζ^*.

Proof. Similarly to the proof of Lemma 10, we apply Kallenberg's theorem, see [68, Proposition 3.22]. Since ζ^* is diffuse, to prove Lemma 18, it is enough to show that, for every precompact relatively open box $B \subset [-\infty, \infty) \times [0, \infty) \times [0, \infty]$, we have

(a) $\mathbb{P}(\Psi_t^*(B) = 0) \to \exp(-\zeta^*(B))$, as $t \uparrow \infty$, and

(b) $\mathbb{E}[\Psi_t^*(B)] \to \zeta^*(B)$, as $t \uparrow \infty$.

It suffices to consider nonempty boxes B of the form $(s_0, s_1) \times (f_0, f_1) \times (z_0, z_1)$, since almost surely, neither the point process $\hat{\Gamma}_t$ nor the limiting Poisson process put points on the boundary $\partial([-\infty, \infty) \times [0, \infty) \times [0, \infty])$. Note that $s_0 = -\infty$ and $z_1 = \infty$ is an allowed choice; and we have

$$\zeta^*(B) = \big(\mathrm{e}^{\lambda s_1} - \mathrm{e}^{\lambda s_0}\big)\big(f_1^\alpha - f_0^\alpha\big) \int_{z_0}^{z_1} \nu(x)\mathrm{d}x.$$

(a) By assumption, $(F_n, \xi_n)_{n\geq 1}$ is a sequence of i.i.d. random variables with each F_n being independent of ξ_n. Hence,

$$\begin{aligned}\mathbb{P}(\Psi_t^*(B)=0) &= \prod_{n(t)\mathrm{e}^{\lambda s_0}<n<n(t)\mathrm{e}^{\lambda s_1}} \mathbb{P}\big(t(1-F_n)\notin (f_0,f_1) \text{ or } \xi_n \notin (z_0,z_1)\big) \\ &= \Big(1-\mathbb{P}\big(t(1-F_1)\in(f_0,f_1)\big)\mathbb{P}\big(\xi_1\in(z_0,z_1)\big)\Big)^{r_{s_0,s_1}(t)},\end{aligned}$$

where $r_{s_0,s_1(t)}$ denotes the number of elements $n\in\mathbb{N}$, with $n(t)\mathrm{e}^{\lambda s_0}<n<n(t)\mathrm{e}^{\lambda s_1}$. We note that, as $t\to\infty$, we have $r_{s_0,s_1}(t)\sim {}^{(\mathrm{e}^{\lambda s_1}-\mathrm{e}^{\lambda s_0})}\!/\!_{\mu(1-t^{-1},1)}$ and, in view of Assumption (B.5),

$$\frac{\mu\big((1-{}^{f_1}\!/\!_t, 1-{}^{f_0}\!/\!_t)\big)}{\mu(1-{}^{1}\!/\!_t,1)}\sim f_1^\alpha-f_0^\alpha \quad \text{as } t\uparrow\infty.$$

Further $\mathbb{P}\big(\xi_1\in(z_0,z_1)\big)=\int_{z_0}^{z_1}\nu(x)\mathrm{d}x$. Thus, as $t\to\infty$

$$\begin{aligned}\mathbb{P}(\Psi_t^*(B)=0) &= \Big(1-\mu\big((1-{}^{f_1}\!/\!_t,1-{}^{f_0}\!/\!_t)\big)\int_{z_0}^{z_1}\nu(x)\mathrm{d}x\Big)^{r_{s_0,s_1}(t)} \\ &\sim \exp\left(-(\mathrm{e}^{\lambda s_1}-\mathrm{e}^{\lambda s_0})\frac{\mu(1-{}^{f_1}\!/\!_t,1-{}^{f_0}\!/\!_t)}{\mu(1-{}^{1}\!/\!_t,1)}\int_{z_0}^{z_1}\nu(x)\,\mathrm{d}x\right) \\ &\to \exp\big(-\zeta^*(B)\big).\end{aligned}$$

(b) To compute the limit of $\mathbb{E}[(\Psi_t^*(B))]$ we apply the asymptotic estimates from above,

$$\begin{aligned}\mathbb{E}[(\Psi_t^*(B))] &= \sum_{n(t)\mathrm{e}^{\lambda s_0}<n<n(t)\mathrm{e}^{\lambda s_1}} \mu\big([1-{}^{f_1}\!/\!_t,1-{}^{f_0}\!/\!_t]\big)\mathbb{P}\big(\xi_n\in[z_0,z_1]\big) \\ &= r_{s_0,s_1}(t)\mu\big([1-{}^{f_1}\!/\!_t,1-{}^{f_0}\!/\!_t]\big)\mathbb{P}\big(\xi_n\in[z_0,z_1]\big)\to\zeta^*(B).\end{aligned}$$

Proof of Lemma 18 is complete. □

Lemma 19. *For any Lipschitz continuous, compactly supported function* $f\colon(-\infty,\infty)\times[0,\infty)\times[0,\infty]\to\mathbb{R}$,

$$\left|\int f\mathrm{d}\Psi_t^*-\int f\mathrm{d}\Psi_t\right|\to 0 \textit{ in probability, as } t\uparrow\infty.$$

Proof. Let f be a Lipschitz function supported on $K=[-a,a]\times[0,b]\times[0,\infty]$ for $a,b\geq 1$.

We have

$$
\begin{aligned}
&\left| \int f \mathrm{d}\Psi_t - \int f \mathrm{d}\Psi_t^* \right| \\
&\quad \le \sum_{n=1}^{M(t)} \left| f\Big(\tau_n - \sigma_t, (t-\tau_n)(1-F_n), \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t)\Big) - f\Big(\tau_n - \sigma_t, t(1-F_n), \xi_n\Big) \right| \\
&\quad \le c_L \sum_{n \in \hat{I}(t)} \Big(\tau_n(1-F_n) + \left| \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \xi_n \right| \Big), \qquad (3.1)
\end{aligned}
$$

where c_L is the Lipschitz constant of the function f and $\hat{I}(t)$ is the random set of indices $n \in \mathbb{N}$ such that

$$|\tau_n - \sigma_t| \le a \quad \text{and} \quad t(1-F_n) \le b.$$

Note that $t(1-F_n) \le b$ implies that $(t-\tau_n)(1-F_n) \le b$. Assume t is large. For $n \in \hat{I}(t)$, $|\tau_n - \sigma_t| \le a$ implies

$$\sigma_t - a \le \tau_n \le \sigma_t + a.$$

Therefore we have

$$\tau_n(1-F_n) \le (\sigma_t + a)\frac{b}{t}.$$

For $\varepsilon \in (0, 1/2)$ define the event $\Upsilon_\varepsilon(t)$ that

$$\Upsilon_\varepsilon(t) := \{\forall n \in \mathbb{N},\ |T + \varepsilon_n| \le \varepsilon\sqrt{\sigma_t}\}.$$

Note that by Assumption (A.1), and Lemma (1.8), $\mathbb{P}(\Upsilon_\varepsilon(t)) \to 1$ as $t \to 1$ for all $\varepsilon > 0$. Let

$$\bar{I}(t) := \{n \in \mathbb{N} : |\tau_n^* - \sigma_t| \le 2|T| + a \text{ and } t(1-F_n) \le b\}.$$

Assume $\hat{I}(t)$ holds on $\Upsilon_\varepsilon(t)$, then

$$|\tau_n^* - \sigma_t| \le |\tau_n^* - \tau_n| + |\tau_n - \sigma_t| \le |\tau_n^* - \tau_n^* - T - \varepsilon_n| + a \le |T| + |\varepsilon_n| + a \le 2|T| + a,$$

which implies that $\bar{I}(t)$ also holds. Hence $\bar{I}(t) \subseteq I_\kappa(t)$ (as defined in (1.9)), on $\Upsilon_\varepsilon(t)$.

Recall the definition of $\Delta_n(t)$ from Assumption (A.2), let $\xi_n := \lim_{t\to\infty} \mathrm{e}^{-\gamma t} Y_n(t)$ and define

$$R_n(t) := \sup_{w > t} |\mathrm{e}^{-\gamma w} Y_n(w) - \xi_n|.$$

By Assumption (A.3) we have $R_n(t) \to 0$ in probability as $t \to \infty$ and, using that for

$t \geq 2b$ and $n \in \hat{I}(t)$ one has $F_n \geq 1/2$ we conclude that for large t,

$$\begin{aligned}
\left| \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \xi_n \right| &\leq \left| \mathrm{e}^{-\gamma F_n(t-\tau_n)} Z_n(t) - \mathrm{e}^{-\gamma F_n(t-\tau_n)} Y_n(F_n(t-\tau_n)) \right| \\
&\quad + \left| \mathrm{e}^{-\gamma F_n(t-\tau_n)} Y_n(F_n(t-\tau_n)) - \xi_n \right| \\
&\leq \Delta_n(F_n(t-\tau_n)) + R_n(F_n(t-\tau_n)) \\
&\leq \Delta_n\left(\tfrac{2t}{3}\right) + R_n\left(\tfrac{2t}{3}\right).
\end{aligned}$$

Therefore, for sufficiently large t, we get

$$\left| \int f \mathrm{d}\Psi_t - \int f \mathrm{d}\Psi_t^* \right| \leq c_L |\hat{I}(t)| (\sigma_t + a) \frac{b}{t} + c_L \sum_{n \in \hat{I}(t)} \Delta_n\left(\tfrac{2t}{3}\right) + c_L \sum_{n \in \hat{I}(t)} R_n\left(\tfrac{2t}{3}\right).$$

Recall that $\lim_{t\to\infty} \sigma_t/t = 0$, by definition of σ_t (see Equation (1.8)). Furthermore, by Lemma 18, $|\hat{I}(t)|$ converges in distribution to an almost surely finite random variable. This implies that

$$\lim_{t\to\infty} c_L |\hat{I}(t)| (\sigma_t + a) \frac{b}{t} = 0, \quad \text{in probability.}$$

By assumption, the random processes $(R_n)_{n\geq 1}$ are independent of $(F_n)_{n\geq 1}$ and thus also of the random set $\hat{I}(t)$. Recall that, by Assumption (A.3)

$$\lim_{v\to\infty} R_n(v) = 0, \quad \text{in probability.}$$

Combining this with Lemma 18, implies that

$$\lim_{t\to\infty} \sum_{n \in \hat{I}(t)} R_n\left(\tfrac{2t}{3}\right) = 0, \quad \text{in probability.}$$

To prove that $\sum_{n\in\hat{I}(t)} \Delta_n\left(\frac{2t}{3}\right) \to 0$ in probability as $t \to \infty$ we use Assumption (A.2). Denoting with $\mathbf{F} = (F_m)_{m\in\mathbb{N}}$, we have

$$\begin{aligned}
\mathbb{P}\Big(\sum_{n\in\hat{I}(t)} \Delta_n\left(\tfrac{2t}{3}\right) \geq \varepsilon \Big) &= \mathbb{E}\Big[\mathbb{P}\Big(\sum_{n\in\hat{I}(t)} \Delta_n\left(\tfrac{2t}{3}\right) \geq \varepsilon \Big| \mathbf{F} \Big) \Big] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}\Big[\mathbb{P}\Big(\exists n \in \hat{I}(t) : \Delta_n\left(\tfrac{2t}{3}\right) \geq \frac{\varepsilon}{k} \Big| \mathbf{F} \Big) \mathbf{1}_{\{|\hat{I}(t)|=k\}} \Big] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}\Big[\sum_{n\in\hat{I}(t)} \mathbb{P}\Big(\Delta_n\left(\tfrac{2t}{3}\right) \geq \frac{\varepsilon}{k} \Big| \mathbf{F} \Big) \mathbf{1}_{\{|\hat{I}(t)|=k\}} \Big] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}\Big[k \max_{n\in\hat{I}(t)} \mathbb{P}\Big(\Delta_n\left(\tfrac{2t}{3}\right) \geq \frac{\varepsilon}{k} \Big| \mathbf{F} \Big) \mathbf{1}_{\{|\hat{I}(t)|=k\}} \Big].
\end{aligned}$$

Now, given $\delta > 0$ pick $K \in \mathbb{N}$ such that, for sufficiently large t,

$$\sum_{k=K+1}^{\infty} \mathbb{E}\Big[k \max_{n \in \hat{I}(t)} \mathbb{P}\Big(\Delta_n(\tfrac{2t}{3}) \geq \frac{\varepsilon}{k}\Big| \mathbf{F}\Big) \mathbf{1}_{\{|\hat{I}(t)|=k\}}\Big] \leq \mathbb{E}\Big[|\hat{I}(t)| \mathbf{1}_{\{|\hat{I}(t)|>K\}}\Big] \leq \frac{\delta}{2},$$

where we use that by Lemma 18 $|\hat{I}(t)|$ converges in distribution to a Poisson random variable and so $\mathbb{E}|\hat{I}(t)|$ converges to its parameter. Given K and using that $\hat{I}(t) \subseteq I_\kappa(t)$ on $\Upsilon_\varepsilon(t)$, we get

$$\sum_{k=0}^{K} \mathbb{E}\Big[k \max_{n \in \hat{I}(t)} \mathbb{P}\Big(\Delta_n(\tfrac{2t}{3}) \geq \frac{\varepsilon}{k}\Big| \mathbf{F}\Big) \mathbf{1}_{\{|\hat{I}(t)|=k\}}\Big] \leq K(K+1)\mathbb{E}\Big[\max_{n \in I_\kappa(t)} \mathbb{P}\Big(\Delta_n(\tfrac{2t}{3}) \geq \frac{\varepsilon}{K}\Big| \mathbf{F}\Big)\Big],$$

which converges to zero by (A.2) and dominated convergence.

This shows that $\sum_{n \in \hat{I}(t)} \Delta_n(\frac{2t}{3}) \to 0$ in probability as $t \to \infty$, which concludes the proof. $\square$

Proof of Proposition 17. Let $f : \mathbb{R} \times [0, \infty) \times [0, \infty] \to \mathbb{R}$ be Lipschitz continuous and compactly supported. Combining Lemmas 18 and 19, together with Slutsky's theorem (see for example [42, Chapter 7.2]) we get the desired result,

$$\int f \mathrm{d}\Psi_t \Rightarrow \int f \mathrm{d}\mathrm{PPP}(\zeta^*) \quad \text{as } t \uparrow \infty,$$

where $\mathrm{PPP}(\zeta^*)$ denotes the Poisson point process with intensity ζ^*. $\square$

3.1.2 n-independence of the second coordinate of Γ_t

Building on the results from the previous subsection, we prove the convergence of Γ_t. Once we show that Γ_t and $\hat{\Gamma}_t$ are asymptotically equivalent, the convergence of Γ_t follows by Proposition 16.

Lemma 20. *For all Lipschitz-continuous compactly-supported functions $f \colon (-\infty, \infty) \times [0, \infty) \times [0, \infty] \to \mathbb{R}$,*

$$\Big| \int f \, \mathrm{d}\Gamma_t - \int f \, \mathrm{d}\hat{\Gamma}_t \Big| \to 0 \quad \textit{in probability as } t \uparrow \infty.$$

Proof. Let f be a Lipschitz function with Lipschitz constant c_L supported on $K = [-a, a] \times [0, b] \times [0, \infty]$ for $a, b \geq 1$.

We have

$$\begin{aligned}
&\Big|\int f \,\mathrm{d}\Gamma_t - \int f \,\mathrm{d}\hat{\Gamma}_t\Big| \\
&\qquad \le \sum_{n=1}^{M(t)} \Big| f\Big(\tau_n - \sigma_t, t(1-F_n), \mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t)\Big) \\
&\qquad\qquad\qquad - f\Big(\tau_n - \sigma_t, (t-\tau_n)(1-F_n), \mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t)\Big)\Big| \\
&\qquad \le c_L \sum_{n\in\hat{I}(t)} \tau_n(1-F_n), \qquad (3.2)
\end{aligned}$$

where $\hat{I}(t)$ is the random set of indices $n \in \mathbb{N}$ such that

$$|\tau_n - \sigma_t| \le a \quad \text{and} \quad t(1-F_n) \le b.$$

We can bound the right-hand side of Equation (3.2) as follows:

$$\begin{aligned}
c_L \sum_{n\in\hat{I}(t)} \tau_n(1-F_n) &\le c_L \sum_{n\in\hat{I}(t)} (\sigma_t + a)\frac{b}{t} \\
&\le \frac{c_L b(\sigma_t + a)}{t} |\hat{I}(t)| \to 0 \quad \text{as } t\to\infty,
\end{aligned}$$

since $\lim_{t\to\infty} \sigma_t/t = 0$, by definition of σ_t (see Equation (1.8)) and $|\hat{I}(t)|$ is almost surely finite since it converges in distribution to a Poisson random variable by Lemma 18. □

Proposition 21. *The point process Γ_t converges vaguely in distribution on $(-\infty,\infty)\times[0,\infty)\times[0,\infty]$ to the Poisson point process with intensity*

$$\mathrm{d}\zeta(s,f,z) = \alpha f^{\alpha-1}\lambda \mathrm{e}^{\lambda s}\mathrm{e}^{\gamma(s+f)}\nu\big(z\,\mathrm{e}^{\gamma(s+f)}\big)\mathrm{d}s\,\mathrm{d}f\,\mathrm{d}z.$$

Proof. By Lemma 20 the point process Γ_t is asymptotically equivalent to $\hat{\Gamma}_t$. Furthermore, by Proposition 16 the point process $\hat{\Gamma}_t$ converges vaguely in distribution on $(-\infty,\infty)\times[0,\infty)\times[0,\infty]$ to the Poisson point process with intensity $\mathrm{d}\zeta$, which implies the result. □

3.2 Negligibility of families outside the main window

Analogously to the proofs in Section 2.2, to deduce Theorem 1 from Proposition 21, one has to control the contribution of the point process near the closed boundaries of $[-\infty,\infty]\times[0,\infty]\times(0,\infty]$. We prove that families that are born too late (Section 3.2.2) or not fit enough (Section 3.2.1), even if they are born early (Section 3.2.3), are too small to contribute in the limit. They get absorbed by the open lower bound of the third coordinate. Our proofs rely on the following result. Recall that

$$n(t) = \left\lceil \frac{1}{\mu(1-t^{-1},1)} \right\rceil = \frac{t^\alpha}{\ell(t^{-1})}.$$

Lemma 22. *Let F be a random variable with law μ. There exists $t_0 > 0$ such that, for all $C \geq 0$, $D > 0$, there exists $K = K(C,D) > 0$ such that*

$$\mathbb{E}\Big[\mathbf{1}\{F \leq 1 - {}^{C}\!/_{t}\}\mathrm{e}^{-D(1-F)t}\Big] \leq \frac{K}{n(t)} \quad \textit{for all } t \geq t_0.$$

Moreover, for all D, we have $\lim_{C\uparrow\infty} K(C,D) = 0$.

For a proof of Lemma 22 see [26, Section 6].

3.2.1 Contribution of the unfit families

Lemma 23. *For every $\eta > 0$ and $c > 0$ there exists $\kappa > 0$ such that, for all sufficiently large t, we have*

$$\mathbb{P}\Big(\max_{n \leq M(t)} \mathbf{1}_{\{F_n \leq 1-\kappa/t\}} Z_n(t) \geq c\mathrm{e}^{\gamma(t-\sigma_t)}\Big) \leq \eta.$$

Proof. Let $c > 0$ and $\kappa > 0$. We analyse the event that there exists a family with fitness $F_n \leq 1 - \kappa/t$ and size $\mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t) \geq c$.

For all $n \geq 1$, we define

$$A_n := \max_{u \geq \tau_n} Z_n(u)\mathrm{e}^{-\gamma F_n(u-\tau_n)}. \tag{3.3}$$

Further, we can rearrange the expression by adding and subtracting $\gamma F_n \sigma_t$ in the exponent

$$\begin{aligned}
&\Big\{Z_n(t) \geq c\mathrm{e}^{\gamma(t-\sigma_t)}\Big\} \\
&\quad\subset \Big\{A_n \geq c\mathrm{e}^{\gamma(t-\sigma_t)-\gamma F_n(t-\tau_n)}\Big\} = \Big\{A_n \geq c\mathrm{e}^{\gamma(1-F_n)(t-\sigma_t)-\gamma F_n(\sigma_t-\tau_n)}\Big\} \\
&\quad= \Big\{A_n \geq c\mathrm{e}^{\gamma(1-F_n)(t-\frac{1}{\lambda}\log n(t)-T-\varepsilon_{n(t)})-\gamma F_n(\frac{1}{\lambda}\log n(t)+T+\varepsilon_{n(t)}-\frac{1}{\lambda}\log n-T-\varepsilon_n)}\Big\} \\
&\quad= \Big\{A_n \geq c\mathrm{e}^{\gamma(1-F_n)(t-\frac{1}{\lambda}\log n(t))-\gamma F_n(\frac{1}{\lambda}\log n(t)-\frac{1}{\lambda}\log n)-\gamma(1-F_n)T+\gamma F_n\varepsilon_n-\gamma\varepsilon_{n(t)}}\Big\},
\end{aligned}$$

where we recalled the definitions of $\tau_n = \frac{1}{\lambda}\log n + T + \varepsilon_n$ (Assumption (A.1)) and $\sigma_t = \tau_{n(t)}$ (Equation (1.8)). Denote by

$$c_{n,t} := c\mathrm{e}^{\gamma(1-F_n)(t-\frac{1}{\lambda}\log n(t))-\gamma F_n(\frac{1}{\lambda}\log n(t)-\frac{1}{\lambda}\log n)}. \tag{3.4}$$

Using the fact that for any random variable Y_n and $c \in \mathbb{R}$ we have $\{\max_{n\in\mathbb{N}} Y_n \geq c\} =$

$\bigcup_{n\in\mathbb{N}}\{Y_n \geq c\}$, we get

$$\mathbb{P}\Big(\max_{n\leq M(t)} \mathbf{1}_{\{F_n\leq 1-\kappa/t\}} Z_n(t) \geq c\,\mathrm{e}^{\gamma(t-\sigma_t)}\Big)$$
$$\leq \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \Big\{\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} A_n \geq c_{n,t}\mathrm{e}^{-\gamma(1-F_n)T+\gamma F_n\varepsilon_n-\gamma\varepsilon_{n(t)}}\Big\}\Big). \qquad (3.5)$$

Furthermore, for any $y_1, y_2 \in \mathbb{R}$, we have

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \Big\{\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} A_n \geq c_{n,t}\mathrm{e}^{-\gamma(1-F_n)T+\gamma F_n\varepsilon_n-\gamma\varepsilon_{n(t)}}\Big\}\Big)$$
$$\leq \mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \Big\{\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} A_n \geq c_{n,t}\mathrm{e}^{-\gamma(1-F_n)y_1-\gamma(F_n+1)y_2}\Big\}\Big)$$
$$+\mathbb{P}(|T| \geq y_1) + \mathbb{P}(\sup_{n\in\mathbb{N}} |\varepsilon_n| \geq y_2). \qquad (3.6)$$

For $y = y_1 = y_2 > 0$, this gives

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} \Big\{\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} A_n \geq c_{n,t}\mathrm{e}^{-\gamma(1-F_n)T+\gamma F_n\varepsilon_n-\gamma\varepsilon_{n(t)}}\Big\}\Big)$$
$$\leq \sum_{n\in\mathbb{N}} \mathbb{P}\Big(\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} A_n \geq c_{n,t}\mathrm{e}^{-2\gamma y}\Big) + \mathbb{P}(|T| \geq y) + \mathbb{P}(\sup_{n\in\mathbb{N}} |\varepsilon_n| \geq y).$$

Since $|\varepsilon_n|$ is bounded and $|T|$ is finite, we can fix $y > 0$ large enough, such that $\mathbb{P}(|T| \geq y) \leq \frac{\eta}{4}$ and $\mathbb{P}(\sup_{n\in\mathbb{N}} |\varepsilon_n| \geq y) \leq \frac{\eta}{4}$.

Consider

$$\begin{aligned} S &:= \sum_{n\in\mathbb{N}} \mathbb{P}\Big(\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} A_n \geq c_{n,t}\mathrm{e}^{-2\gamma y}\Big) \\ &= \sum_{n\in\mathbb{N}} \mathbb{E}\Big[\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} \mathbb{P}\Big(A_n \geq c_{n,t}\mathrm{e}^{-2\gamma y} | (F_m)_{m\in\mathbb{N}}\Big)\Big]. \end{aligned}$$

By Assumption (A.4), $\mathbb{P}(A_n \geq u | (F_m)_{m\in\mathbb{N}}) \leq c_0\mathrm{e}^{-\eta u}$, which implies

$$\begin{aligned} S &\leq c_0 \sum_{n\in\mathbb{N}} \mathbb{E}\Big[\mathbf{1}_{\{F_n\leq 1-\kappa/t\}} \exp\Big\{-\eta c\mathrm{e}^{\gamma(1-F_n)(t-\frac{1}{\lambda}\log n(t))-\gamma F_n(\frac{1}{\lambda}\log n(t)-\frac{1}{\lambda}\log n)-2\gamma y}\Big\}\Big] \\ &\leq c_0 \sum_{n\in\mathbb{N}} \mathbb{E}\Big[\mathbf{1}_{\{F\leq 1-\kappa/t\}} \exp\Big\{-\eta c\mathrm{e}^{\gamma(1-F)t/2-\gamma F/\lambda \log(n(t)/n)-2\gamma y}\Big\}\Big] \end{aligned}$$

where F is a random variable of law μ, and t is large so that $t/2 \geq \frac{1}{\lambda}\log n(t)$ (this is guaranteed since $\frac{1}{\lambda}\log n(t) = \frac{\alpha}{\lambda}\log t - \frac{1}{\lambda}\log \ell(t^{-1})$ which is deterministic).

We now fix small numbers $\delta, \rho > 0$ and note that there exists a constant C_ρ such that

$e^{-x} \le C_\rho x^{-\rho}$ for all $x \ge 0$, so we can bound

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\{\delta \le F \le 1-\kappa/t\}} \exp\Big\{ -\eta c e^{\gamma(1-F)t/2 - \gamma F/\lambda \log(n(t)/n) - 2\gamma y}\Big\}\Big] \\
&\qquad \le \mathbb{E}\Big[\mathbf{1}_{\{\delta \le F \le 1-\kappa/t\}} C_\rho \Big(\eta c e^{\gamma(1-F)t/2 - \gamma F/\lambda \log(n(t)/n) - 2\gamma y}\Big)^{-\rho}\Big] \\
&\qquad \le C_\rho (\eta c)^{-\rho} e^{2\gamma\rho y} \mathbb{E}\Big[\mathbf{1}_{\{\delta \le F \le 1-\kappa/t\}} e^{-\gamma\rho(1-F)t/2} \Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\rho F}{\lambda}}\Big]
\end{aligned}$$

For $n \ge n(t)$, we can replace F by δ to get

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\{\delta \le F \le 1-\kappa/t\}} \exp\Big\{ -\eta c e^{\gamma(1-F)t/2 - \gamma F/\lambda \log(n(t)/n) - 2\gamma y}\Big\}\Big] \\
&\qquad \le C_\rho (\eta c)^{-\rho} e^{2\gamma\rho y} \Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\rho\delta}{\lambda}} \mathbb{E}\Big[\mathbf{1}_{\{F \le 1-\kappa/t\}} e^{-\gamma\rho(1-F)t/2}\Big] \\
&\qquad \le C_\rho (\eta c)^{-\rho} e^{2\gamma\rho y} \Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\rho\delta}{\lambda}} \frac{K(\kappa, \rho\gamma/2)}{n(t)}, \qquad (3.7)
\end{aligned}$$

where we used Lemma 22. Similarly, for $n \le n(t)$, we replace F by 1, to get

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\{\delta \le F \le 1-\kappa/t\}} \exp\Big\{ -\eta c\, e^{\gamma(1-F)t/2 - \gamma F/\lambda \log(n(t)/n) - 2\gamma y}\Big\}\Big] \\
&\qquad\qquad \le C_\rho (\eta c)^{-\rho} e^{2\gamma\rho y} \Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\rho}{\lambda}} \frac{K(\kappa, \rho\gamma/2)}{n(t)}. \qquad (3.8)
\end{aligned}$$

Applying (3.7) with $\rho_+ > \frac{\lambda}{\gamma\delta}$, if $n > n(t)$, and (3.8) with $\rho_- < \frac{\lambda}{\gamma}$, if $n \le n(t)$, we get

$$\begin{aligned}
&\sum_{n\in\mathbb{N}} \mathbb{P}\Big(\mathbf{1}_{\{F_n \le 1-\kappa/t\}} A_n \ge c_{n,t} e^{-2\gamma y}\Big) \\
&\quad \le c_0 C_{\rho_-} (\eta c)^{-\rho_-} e^{2\gamma y \rho_-} \frac{K(\kappa, \gamma\rho_-/2)}{n(t)} \sum_{n=1}^{\lceil n(t)\rceil} \Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\rho_-}{\lambda}} \\
&\qquad\qquad + c_0 C_{\rho_+} (\eta c)^{-\rho_+} e^{2\gamma y \rho_+} \frac{K(\kappa, \gamma\rho_+/2)}{n(t)} \sum_{n=\lceil n(t)\rceil+1}^{\infty} \Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\delta\rho_+}{\lambda}} + c_0\mathbb{P}(F<\delta) \\
&\quad \le C\big(K(\kappa, \gamma\rho_-/2) + K(\kappa, \gamma\rho_+/2)\big) + c_0\mathbb{P}(F<\delta),
\end{aligned}$$

where C is a constant not depending on κ or t, using that both sums are bounded by a constant multiple of $n(t)$. Recalling (3.5) and (3.6), we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{n \ge M(t)} \mathbf{1}_{\{F_n \le 1-\kappa/t\}} Z_n(t) \ge c\, e^{\gamma(t-\sigma_t)}\Big) \\
&\quad \le C\big(K(\kappa, \gamma\rho_-/2) + K(\kappa, \gamma\rho_+/2)\big) + c_0\mathbb{P}(F<\delta) + \mathbb{P}(|T| \ge y) + \mathbb{P}\big(\sup_{n\in\mathbb{N}} |\varepsilon_n| \ge y\big).
\end{aligned}$$

Recalling that $\lim_{\kappa\to\infty} K(\kappa, \gamma\rho_\pm/2) = 0$ and $\mathbb{P}(F < \delta) \to 0$ as $\delta \downarrow 0$ completes the proof. $\square$

3.2.2 Contribution of the families born late

Lemma 24. *For every $\eta > 0$ and $c > 0$ there exists $v > 1$ such that, for all sufficiently large t, we have*

$$\mathbb{P}\Big(\max_{vn(t) \le n \le M(t)} Z_n(t) \ge c\mathrm{e}^{\gamma(t-\sigma_t)} \Big) \le \eta.$$

Proof. The proof is similar to the proof of Lemma 23. Let $c > 0$ and define the sequence (A_n) as in Equation (3.3). We have

$$\begin{aligned}
&\{Z_n(t) \ge c\,\mathrm{e}^{\gamma(t-\sigma_t)}\} \\
&\qquad \subset \Big\{ A_n \ge c\,\mathrm{e}^{\gamma(1-F_n)(t-\sigma_t) - \gamma F_n(\sigma_t - \tau_n)} \Big\} \\
&\qquad = \Big\{ A_n \ge c\mathrm{e}^{\gamma(1-F_n)(t-\frac{1}{\lambda}\log n(t)) - \gamma F_n(\frac{1}{\lambda}\log n(t) - \frac{1}{\lambda}\log n) - \gamma(1-F_n)T + \gamma F_n \varepsilon_n - \gamma\varepsilon_{n(t)}} \Big\} \\
&\qquad = \Big\{ A_n \ge c_{n,t}\mathrm{e}^{-\gamma(1-F_n)T + \gamma F_n\varepsilon_n - \gamma\varepsilon_{n(t)}} \Big\},
\end{aligned}$$

where $c_{n,t}$ is the same as in Equation (3.4). Let $v > 1$ and $y > 0$. By an argument analagous to the proof in Lemma 23 we have

$$\mathbb{P}\Big(\max_{n \ge vn(t)} Z_n(t) \ge c\mathrm{e}^{\gamma(t-\sigma_t)} \Big) \le \sum_{n \ge vn(t)} \mathbb{P}\Big(A_n \ge c_{n,t}\mathrm{e}^{-2\gamma y} \Big) + \mathbb{P}(|T| \ge y) + \mathbb{P}\big(\sup_{n\in\mathbb{N}} |\varepsilon_n| \ge y \big), \tag{3.9}$$

where we can fix $y > 0$ to be large so that $\mathbb{P}(|T| \ge y) \le \eta/4$ and $\mathbb{P}(\sup_{n\in\mathbb{N}} |\varepsilon_n| \ge y) \le \eta/4$. Furthermore, for t large, so that $t/2 \ge 1/\lambda \log n(t)$, we have

$$\begin{aligned}
\mathbb{P}\Big(A_n \ge c_{n,t}\mathrm{e}^{-2\gamma y} \Big) &= \mathbb{P}\Big(A_n \ge c\mathrm{e}^{\gamma(1-F_n)(t-\frac{1}{\lambda}\log n(t)) - \gamma F_n(\frac{1}{\lambda}\log n(t) - \frac{1}{\lambda}\log n) - 2\gamma y} \Big) \\
&\le \mathbb{P}\Big(A_n \ge c\mathrm{e}^{\gamma(1-F_n)t/2 - \gamma F_n(\frac{1}{\lambda}\log n(t) - \frac{1}{\lambda}\log n) - 2\gamma y} \Big).
\end{aligned}$$

An argument analogous to Lemma 23 yields, for any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(A_n \ge c\mathrm{e}^{\gamma(1-F_n)t/2 - \gamma F_n(\frac{1}{\lambda}\log n(t) - \frac{1}{\lambda}\log n) - 2\gamma y} \Big) \\
&\qquad \le c_0 \mathbb{E}\Big[\exp\Big\{ -\eta c\,\mathrm{e}^{\gamma(1-F)t/2 - \gamma\delta(\frac{1}{\lambda}\log n(t) - \frac{1}{\lambda}\log n) - 2\gamma y} \Big\} \Big] + c_0\,\mathbb{P}(F < \delta),
\end{aligned}$$

where F is a random variable of law μ. We now pick $\rho > \frac{\lambda}{\gamma\delta}$. As in Lemma 23 we use existence of a constant C_ρ such that $\mathrm{e}^{-x} \le C_\rho x^{-\rho}$, for all $x \ge 0$, and Lemma 22 to get

$$\mathbb{E}\Big[\exp\Big\{ -\eta c\,\mathrm{e}^{\gamma(1-F)t/2 - \gamma\delta(\frac{1}{\lambda}\log n(t) - \frac{1}{\lambda}\log n) - 2\gamma y} \Big\} \Big] \le C_\rho(\eta c)^{-\rho}\mathrm{e}^{2\gamma\rho y} \Big(\frac{n(t)}{n} \Big)^{\frac{\gamma\rho\delta}{\lambda}} \frac{K(0, \rho\gamma/2)}{n(t)}.$$

Summing over $n \geq vn(t)$ yields

$$\sum_{n \geq vn(t)} \mathbb{P}\Big(A_n \geq c_{n,t}\mathrm{e}^{-2\gamma y}\Big) \leq c_0 C_\rho(\eta c)^{-\rho}\mathrm{e}^{2\gamma\rho y}\frac{K(0,\rho\gamma/2)}{n(t)}\sum_{n=vn(t)}^{\infty}\Big(\frac{n(t)}{n}\Big)^{\frac{\gamma\rho\delta}{\lambda}} + c_0\,\mathbb{P}(F<\delta)$$
$$\leq Cv^{1-\frac{\gamma\rho\delta}{\lambda}} + c_0\,\mathbb{P}(F<\delta), \tag{3.10}$$

where C is a constant that does not depend on t or v. We have $1-\frac{\gamma\rho\delta}{\lambda} < 0$, and $\mathbb{P}(F<\delta) \to 0$, as $\delta \downarrow 0$. Combining these facts and plugging (3.10) into (3.9) completes the proof. □

3.2.3 Families born early are not fit enough

Lemma 25. *For all $\kappa, \eta > 0$, there exists $w = w(\kappa,\eta) > 0$ such that, for all t large enough,*

$$\mathbb{P}\Big(\Gamma_t([-\infty,-\log w]\times[0,\kappa]\times[0,\infty]) = 0\Big) \geq 1-\eta.$$

Proof. We have

$$\begin{aligned}\mathbb{P}\Big(\Gamma_t([-\infty,-\log w]\times[0,\kappa]\times[0,\infty]) = 0\Big) &= \mathbb{P}(F_n < 1-\kappa/t, \forall n \text{ s.t. } \tau_n \leq \sigma_t - \log w)\\ &= \mathbb{P}(F_n < 1-\kappa/t, \forall n \leq M(\sigma_t-\log w)),\end{aligned}$$

where we recall that $M(\sigma_t - \log w)$ is the number of families that were founded before time $\sigma_t - \log w$.

Hence, by Assumption (B.5), we have

$$\begin{aligned}&\mathbb{P}\Big(\Gamma_t([-\infty,-\log w]\times[0,\kappa]\times[0,\infty]) = 0\Big)\\ &\quad= \mathbb{E}[\mu(0,1-\kappa/t)^{M(\sigma_t-\log w)}] = (1+o(1))\mathbb{E}\Big[\exp\big\{-M(\sigma_t-\log w)(\kappa/t)^{\alpha}\ell(\kappa/t)\big\}\Big],\end{aligned}$$

as t goes to infinity. Note that, in view of Assumption (A.1),

$$\log w = \sigma_t - (\sigma_t - \log w) \leq \tau_{n(t)} - \tau_{M(\sigma_t-\log w)} \leq \frac{1}{\lambda}\log\frac{n(t)}{M(\sigma_t-\log w)},$$

with probability tending to one, implying that

$$M(\sigma_t - \log w) \leq n(t)\mathrm{e}^{-\lambda \log w}.$$

Recalling that, by Assumption (B.5) $n(t) \sim \frac{t^\alpha}{\ell(1/t)}$, we get

$$\mathbb{P}\Big(\Gamma_t([-\infty,-\log w]\times[0,\kappa]\times[0,\infty]) = 0\Big) \geq (1+o(1))\exp\Big\{-\mathrm{e}^{-\lambda\log w}\kappa^{\alpha}\frac{\ell(\kappa/t)}{\ell(1/t)}\Big\}.$$

Since ℓ is a slowly varying function, we have that $\frac{\ell(\kappa/t)}{\ell(1/t)} \to 1$ when t tends to infinity. In conclusion,

$$\mathbb{P}\Big(\Gamma_t([-\infty, -\log w] \times [0,\kappa] \times [0,\infty]) = 0\Big) \geq (1+o(1)) \exp\Big\{ -\mathrm{e}^{-\lambda \log w} \kappa^{\alpha}\Big\} \to 1,$$

as $w \uparrow \infty$, which completes the proof. $\square$

3.3 Completion of proofs

3.3.1 Proof of Theorem 1

Let $\eta, c > 0$. By Lemma 23, there exists $\kappa = \kappa(c,\eta)$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t([-\infty,\infty] \times [\kappa,\infty] \times (c,\infty]) = 0\Big) \geq 1-\eta.$$

By Lemma 24, there exists $v = v(c,\eta) > 1$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t([\log v,\infty] \times [0,\infty] \times (c,\infty]) = 0\Big) \geq 1-\eta.$$

By Lemma 25, there exists $w = w(\kappa,\eta) > 0$ such that

$$\liminf_{t\to\infty} \mathbb{P}\Big(\Gamma_t([-\infty,-\log w] \times [0,\kappa] \times [0,\infty]) = 0\Big) \geq 1-\eta.$$

By Proposition 21, Γ_t converges on $(-\infty, \log v) \times [0,\kappa) \times (c,\infty]$ to the Poisson process with intensity measure ζ. Combining these four facts and using that $\eta > 0$ is arbitrarily small, we get convergence on $[-\infty,\infty] \times [0,\infty] \times (c,\infty]$ which holds for all $c > 0$, and so the proof is complete.

3.3.2 Proof of Corollary 2

Proof of Corollary 2. (i) We fix $x > 0$ and apply the vague convergence proved in Theorem 1 to the compact set $K := [-\infty,\infty] \times [0,\infty] \times [x,\infty]$. As $t \to \infty$, we get

$$\sum_{n=1}^{M(t)} \mathbf{1}_K\big(\tau_n - \sigma_t, t(1-F_n), \mathrm{e}^{-\gamma(t-\sigma_t)} Z_n(t)\big) \Rightarrow \mathrm{Poisson}\Big(\int_K \mathrm{d}\zeta\Big).$$

Therefore we have

$$\begin{aligned}\mathbb{P}\Big(\mathrm{e}^{-\gamma(t-\sigma_t)} \max_{n\in\{1,\dots,M(t)\}} Z_n(t) \geq x\Big) &\to \mathbb{P}\Big(\mathrm{Poisson}\Big(\int_K \mathrm{d}\zeta\Big) \geq 1\Big)\\ &= 1 - \exp\Big(-\int_K \mathrm{d}\zeta\Big).\end{aligned} \tag{3.11}$$

Integrating out gives us

$$\begin{aligned}
\int_K \mathrm{d}\zeta &= \int_{-\infty}^{\infty}\int_0^{\infty}\int_x^{\infty} \alpha f^{\alpha-1}\lambda \mathrm{e}^{\lambda s}\mathrm{e}^{\gamma(s+f)}\nu\big(z\mathrm{e}^{\gamma(s+f)}\big)\,\mathrm{d}z\,\mathrm{d}f\,\mathrm{d}s\\
&= \int_0^{\infty}\nu(w)\int_0^{\infty}\alpha f^{\alpha-1}\int_{-\infty}^{\frac{1}{\gamma}\log\frac{w}{x}-f}\lambda\mathrm{e}^{\lambda s}\,\mathrm{d}s\,\mathrm{d}f\,\mathrm{d}w\\
&= \Big(\int_0^{\infty}\nu(w)\Big(\frac{w}{x}\Big)^{\frac{\lambda}{\gamma}}\,\mathrm{d}w\Big)\Big(\int_0^{\infty}\alpha f^{\alpha-1}\mathrm{e}^{-\lambda f}\,\mathrm{d}f\Big)\\
&= \frac{\Gamma(\alpha+1)}{\lambda^{\alpha}}\Big(\int_0^{\infty}\nu(w)w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w\Big)x^{-\frac{\lambda}{\gamma}}.
\end{aligned}$$

Thus, the right hand side in (3.11) is $1-\exp\big(-s^{\eta}x^{-\eta}\big)$, for

$$s^{\eta} = \Gamma(\alpha+1)\lambda^{-\alpha}\int_0^{\infty}\nu(w)w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w, \quad \text{and } \eta = \frac{\lambda}{\gamma}.$$

In summary, for all $y>0$, we have

$$\mathbb{P}\Big(\mathrm{e}^{-\gamma(t-\sigma_t)}\max_{n\in\{1,\dots,M(t)\}} Z_n(t)\le y\Big)\to \exp\Big\{-\Big(\frac{y}{s}\Big)^{-\frac{\lambda}{\gamma}}\Big\} = \mathbb{P}(W\le y),$$

as $t\to\infty$, where $W\sim$ Fréchet$\big(\frac{\lambda}{\gamma},s\big)$. Putting this together with Assumption (B.5) and Equation (1.8) gives us the desired result.

(ii) By Theorem 1 the probability that the random variable $t(1-V(t))$ is in an interval $[a,b]$ for some $a<b$, converges to

$$\int_{-\infty}^{\infty}\int_a^b\int_0^{\infty}\mathrm{e}^{-\zeta([-\infty,\infty]\times[0,\infty]\times[z,\infty])}\zeta(\mathrm{d}s,\ \mathrm{d}f,\ \mathrm{d}z).$$

Recall from the proof of part (i), that

$$\zeta([-\infty,\infty]\times[0,\infty]\times[z,\infty]) = \Gamma(\alpha+1)\lambda^{-\alpha}\Big(\int_0^{\infty}\nu(w)w^{\frac{\lambda}{\gamma}}\,\mathrm{d}w\Big)z^{-\frac{\lambda}{\gamma}}.$$

Combining this with the definition of $\zeta(\mathrm{d}s,\ \mathrm{d}f,\ \mathrm{d}z)$ and substituting $v=\mathrm{e}^{\gamma(s+f)}$, we get

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_0^{\infty}\mathrm{e}^{-\zeta([-\infty,\infty]\times[0,\infty]\times[z,\infty])}\zeta(\mathrm{d}s,\ f,\ \mathrm{d}z)\\
&\qquad = \alpha f^{\alpha-1}\mathrm{e}^{-\lambda f}\int_0^{\infty}\Big(\int_0^{\infty}\frac{\lambda}{\gamma}v^{\frac{\lambda}{\gamma}}\nu(zv)\,\mathrm{d}v\Big)\mathrm{e}^{-\Gamma(\alpha+1)\lambda^{-\alpha}\left(\int_0^{\infty}\nu(w)w^{\frac{\lambda}{\gamma}}\mathrm{d}w\right)z^{-\frac{\lambda}{\gamma}}}\,\mathrm{d}z.
\end{aligned}$$

Observing that this is a probability density, of the form $cf^{\alpha-1}\mathrm{e}^{-\lambda f}$, we conclude that

$$c = \alpha\int_0^{\infty}\Big(\int_0^{\infty}\frac{\lambda}{\gamma}v^{\frac{\lambda}{\gamma}}\nu(zv)\,\mathrm{d}v\Big)\mathrm{e}^{-\Gamma(\alpha+1)\lambda^{-\alpha}\left(\int_0^{\infty}\nu(w)w^{\frac{\lambda}{\gamma}}\mathrm{d}w\right)z^{-\frac{\lambda}{\gamma}}}\,\mathrm{d}z = \frac{\lambda^{\alpha}}{\Gamma(\alpha)},$$

and so V is Gamma distributed with shape parameter α and scale parameter λ.

(iii) By Theorem 1 the random variable $S(t) - \sigma_t$ converges to a random variable U with density

$$\int_0^\infty \int_0^\infty \mathrm{e}^{-\zeta([-\infty,\infty]\times[0,\infty]\times[z,\infty])}\zeta(s,\ \mathrm{d}f,\ \mathrm{d}z). \qquad \square$$

Chapter 4

Examples and applications

In this chapter we present a selection of examples covered by our main results. We emphasise that our framework goes well beyond the setup of reinforced branching processes treated in [26] and also that we pick only a small number of representative results out of a wealth of consequences that we can draw from Theorem 3 and Corollary 4. All proofs are done in Section 4.4.

4.1 Branching processes with selection and mutation

We start with one individual with genetic fitness sampled from μ. Individuals never die and give birth at a rate given by their fitness to an independent random number of offspring. Note that variations in individual fitness lead to a selection effect: an individual born at time t selects its parent from the population alive at time t with a probability proportional to their fitness. At birth each individual independently either inherits the parent's fitness or, with probability $0 < \beta < 1$, is a mutant getting a fitness sampled from μ independently of everything else. Similar to the deterministic Kingman's model [48, 28] at mutation all genetic information from a particle's ancestry is lost. For a discussion of the relevance of these models in the theory of evolution see [44].

In our framework the non-decreasing sequence of birth times $\tau_1, \tau_2, \ldots$ of mutants constitute the foundation times of new families, their fitnesses are $F_1, F_2, \ldots$ and $Z_n(t)$ is the number of non-mutant offspring of the nth mutant at time t. If $(p_k)_{k\in\mathbb{N}}$ is the distribution of offspring numbers at a birth event denote by $\mathfrak{m} = \sum k p_k$ the mean offspring number and assume that there exists $\eta > 0$ such that $\sum \mathrm{e}^{\eta k} p_k < \infty$. We assume that mutations have a reasonable chance to produce fit individuals, as expressed in the Malthusian condition

$$\beta \int_0^1 \frac{\mathrm{d}\mu(x)}{1-x} > 1.$$

Under this condition there is a unique solution $\lambda > (1-\beta)\mathfrak{m}$ of the equation

$$\beta\mathfrak{m}\int_0^1 \frac{x}{\lambda-(1-\beta)\mathfrak{m}x}\,\mathrm{d}\mu(x) = 1.$$

We prove in Section 4.4.1 that (A.1) to (A.4) are satisfied with $\gamma = (1-\beta)\mathfrak{m}$. If $p_1 = 1$ this is a reinforced branching process as studied in [26]. The generalisation to arbitrary offspring distribution (p_{ij}) is not difficult, see Section 4.4.1 for details. As an example of the limit theorems we get we look at the birth time $S(t)$ of the largest family at time t in the case of Gnedenko's distribution (Example (3) in Section 1.4.1)

$$\mu(x,1) = \mathrm{e}^{-\frac{x}{1-x}}, \qquad \text{for } 0 < x < 1,$$

see [41, Example 2]. We find a leading order term for $S(t)$ of

$$\sigma_t = \frac{1}{\lambda}(\sqrt{\lambda t+1}-1)$$

and $\varkappa = 2$. Corollary 4 therefore gives a central limit theorem of the form

$$\frac{S(t)-\sqrt{t/\lambda}}{\sqrt[4]{t/\lambda}} \Rightarrow \mathcal{N}(0,(2\lambda)^{-1}).$$

4.2 Preferential attachment networks with fitness

4.2.1 Preferential attachment tree of Bianconi and Barabási

This model is a random tree where at each step a new vertex is added and connected to an existing vertex with a probability depending on the fitness of the vertices. The model was introduced by Bianconi and Barabási in [15] (see Section 1.2 for more details). We start with two vertices connected by an edge, and endowed with fitnesses sampled independently from μ. At every step $n \geq 3$ a new vertex arrives, gets a fitness sampled from μ independently of everything else, and connects to one existing vertex chosen randomly from the $n-1$ existing vertices with a probability proportional to the product of their fitness and their degree.

The preferential attachment tree of Bianconi and Barabási can be embedded in continuous time and then represents a reinforced branching process as in [26], see Section 4.4.1 for details. In this embedding τ_n is the birth time of the nth vertex, F_n its fitness and $Z_n(t)$ its degree at time t. We show in Section 4.4.1 that under the Malthusian condition

$$\int_0^1 \frac{\mu(dx)}{1-x} > 2$$

the process satisfies Assumptions (A.1) to (A.4) with $\gamma = 1$ and $\lambda > 1$ the unique solution of the equation

$$\int_0^1 \frac{x}{\lambda - x} \, \mathrm{d}\mu(x) = 1.$$

We now give an example of our result for the network with fitness distribution

$$\mu(x, 1) = \mathrm{e}^{1-(1-x)^{-\varrho}}, \quad \text{for } 0 < x < 1,$$

where $0 < \varrho < 1$, see Example (1) in Section 1.4.1. We estimate σ_t, as defined in Equation (1.6). Using that $g(x) = m^{-1}(x) = 1 - (x+1)^{-\frac{1}{\varrho}}$, we have that $x = \lambda\sigma_t$ is the unique solution of

$$(\log g)'(x) = \frac{1}{\lambda t + 1 - (x+1)},$$

which we can rewrite as

$$\lambda t + 1 = \varrho(x+1)^{\frac{\varrho+1}{\varrho}} + (1-\varrho)(x+1). \tag{4.1}$$

Let $y = x + 1$. Since $\lambda t + 1 \sim \lambda t$, and the right hand side of Equation (4.1) $\sim \varrho y^{\frac{\varrho+1}{\varrho}}$, we have $y = (\frac{\lambda t}{\varrho})^{\frac{\varrho}{\varrho+1}} + u_t$, where $u_t = o\Big(t^{\frac{\varrho}{\varrho+1}}\Big)$. Rewriting Equation (4.1) gives

$$\lambda t + 1 = \varrho\Big(\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho}{\varrho+1}} + u_t\Big)^{\frac{\varrho+1}{\varrho}} + (1-\varrho)\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho}{\varrho+1}} + (1-\varrho)u_t.$$

Using Taylor's approximation we get

$$1 = u_t(\varrho+1)\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{1}{\varrho+1}}(1+o(1)) + (1-\varrho)\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho}{\varrho+1}} + (1-\varrho)u_t,$$

which implies

$$u_t = -\frac{1-\varrho}{1+\varrho}\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho-1}{\varrho+1}}(1+o(1)).$$

Therefore, we get

$$\lambda\sigma_t = y - 1 = \Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho}{\varrho+1}} - \frac{1-\varrho}{1+\varrho}\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho-1}{\varrho+1}}(1+o(1)). \tag{4.2}$$

This implies

$$\sigma_t = x_0 t^{\frac{\varrho}{\varrho+1}} + \mathcal{O}\Big(t^{\frac{\varrho-1}{\varrho+1}}\Big) \tag{4.3}$$

where $x_0 = \lambda^{-\frac{1}{\varrho+1}} \varrho^{-\frac{\varrho}{\varrho+1}}$.

By Theorem 3, $(\Gamma_t)_{t\geq 0}$ converges vaguely on the space $[-\infty,\infty] \times [-\infty,\infty] \times (0,\infty]$ to the Poisson point process with intensity measure

$$\mathrm{d}\zeta(s,\ f,\ z) = \lambda \mathrm{e}^{-f} \mathrm{e}^{s^2 a_2 - f/\lambda} \exp\{-z\mathrm{e}^{s^2 a_2 - f/\lambda}\} \, \mathrm{d}s \, \mathrm{d}f \, \mathrm{d}z,$$

where $a_2 = \frac{\varrho+1}{2\varrho}$, since $\gamma = 1$, and $\varkappa = \frac{\varrho+1}{\varrho}$ (by Equation (1.12)).

We can now apply Corollary 4. For part (i) we need to substitute $\sigma_t = x_0 t^{\frac{\varrho}{\varrho+1}} + \mathcal{O}\Big(t^{\frac{\varrho-1}{\varrho+1}}\Big)$ into $-\gamma g(\lambda\sigma_t)(t-\sigma_t) - a_1 g(\lambda\sigma_t)\log\sigma_t + \gamma T$. Recall that $g(x) = 1 + (1-x)^{-\frac{1}{\varrho}}$, therefore we can rewrite

$$g(\lambda\sigma_t) = 1 - \bigg(\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho}{\varrho+1}} + \mathcal{O}\Big(t^{\frac{\varrho-1}{\varrho+1}}\Big) + 1\bigg)^{-\frac{1}{\varrho}} = 1 - \Big(\frac{\lambda t}{\varrho}\Big)^{-\frac{1}{\varrho+1}} + \frac{1}{\varrho}\Big(\frac{\lambda t}{\varrho}\Big)^{-1} + \mathcal{O}\Big(t^{\frac{-1-2\varrho}{\varrho+1}}\Big).$$

Furthermore, we have

$$\begin{aligned} &g(\lambda\sigma_t)(t-\sigma_t) \\ &\quad = \bigg(1 - \Big(\frac{\lambda t}{\varrho}\Big)^{-\frac{1}{\varrho+1}} + \frac{1}{\varrho}\Big(\frac{\lambda t}{\varrho}\Big)^{-1} + \mathcal{O}\Big(t^{\frac{-1-2\varrho}{\varrho+1}}\Big)\bigg)\Big(t - \lambda^{-\frac{1}{\varrho+1}}\varrho^{-\frac{\varrho}{\varrho+1}} t^{\frac{\varrho}{\varrho+1}} + \mathcal{O}\Big(t^{\frac{\varrho-1}{\varrho+1}}\Big)\Big). \end{aligned}$$

Simplifying we get

$$g(\lambda\sigma_t)(t-\sigma_t) = t - \frac{1}{\lambda}\Big(\frac{\lambda t}{\varrho}\Big)^{\frac{\varrho}{\varrho+1}} - \Big(\frac{\lambda}{\varrho}\Big)^{-\frac{1}{\varrho+1}} t^{\frac{\varrho}{\varrho+1}} + \frac{1}{\lambda} + \mathcal{O}\Big(t^{-\frac{1}{\varrho+1}}\Big).$$

Denoting by $a_4 := \varrho^{-\frac{\varrho}{\varrho+1}} + \varrho^{\frac{1}{\varrho+1}}$ and $a_5 := \frac{\varrho}{2(\varrho+1)}$, we can now apply Corollary 4(i), to get the following distributional limit for the size of the largest family. Asymptotically as $t \to \infty$,

$$\mathrm{e}^{-\big(t - \lambda^{-\frac{1}{1+\varrho}} a_4 t^{\frac{\varrho}{\varrho+1}} + \frac{1}{\lambda}\big) - \frac{1}{\lambda} a_5 \log t + T} \max_{n\in\mathbb{N}} Z_n(t) \Rightarrow W,$$

where W is a Fréchet distributed random variable with shape parameter λ and scale parameter s given by $s^\lambda = \sqrt{\frac{2\pi\varrho}{\varrho+1}}\Gamma(\lambda+1)$.

To get a result that is independent of the continuous time embedding we look at the time τ_n when the nth vertex is introduced. The largest degree at this instance satisfies[1]

$$\max_{m\le n} Z_m(\tau_n) \asymp n^{\frac{1}{\lambda}} \mathrm{e}^{-\frac{1}{\lambda} a_4 (\log n)^{\frac{\varrho}{\varrho+1}} + \frac{1}{\lambda} a_5 \log\log n},$$

where the implied constants are positive random variables. Recalling that the Bianconi–Barabási tree with bounded fitness has power-law distribution with parameter $\lambda + 1$, we conclude that our result is in line with Equation (1.2).

Corollary 4(ii) implies the following result for the birth time a family of maximal size at time t, $S(t)$. Approximating σ_t by $\sigma_t^* = x_0 t^{\frac{\varrho}{\varrho+1}}$, such that $\sigma_t = \sigma_t^*\Big(1 + \mathcal{O}\Big(t^{-\frac{1}{\varrho+1}}\Big)\Big)$, by Equation (4.3), we can rewrite

$$\frac{S(t) - x_0 t^{\frac{\varrho}{\varrho+1}}}{\sqrt{x_0 t^{\frac{\varrho}{\varrho+1}}}} = \frac{S(t) - \sigma_t}{\sqrt{\sigma_t}} \times \frac{\sqrt{\sigma_t}}{\sqrt{\sigma_t^*}} + \frac{\sigma_t - \sigma_t^*}{\sqrt{\sigma_t^*}} = \frac{S(t) - \sigma_t}{\sqrt{\sigma_t}}\Big(1 + \mathcal{O}\Big(t^{-\frac{1}{\varrho+1}}\Big)\Big) + \mathcal{O}\Big(t^{\frac{\varrho-2}{2(\varrho+1)}}\Big).$$

By Corollary 4(ii) we have $\frac{S(t)-\sigma_t}{\sqrt{\sigma_t}} \Rightarrow U$. So applying Slutsky's theorem (see for exam-

[1] We write $a_n \asymp b_n$ iff $a_n = \mathcal{O}(b_n)$ and $b_n = \mathcal{O}(a_n)$.

ple [42, Chapter 7.2]), we conclude that as $t \to \infty$

$$\frac{S(t) - x_0 t^{\frac{\varrho}{\varrho+1}}}{\sqrt{x_0 t^{\frac{\varrho}{\varrho+1}}}} \Rightarrow U,$$

where U is normally distributed with mean 0 and variance $\frac{\varrho}{\lambda(\varrho+1)}$. This can be formulated for discrete time as follows

$$\frac{S(\tau_n) - \frac{1}{\lambda}\left(\frac{1}{\varrho}\log n\right)^{\frac{\varrho}{\varrho+1}}}{\sqrt{\frac{1}{\lambda}\left(\frac{1}{\varrho}\log n\right)^{\frac{\varrho}{\varrho+1}}}} \Rightarrow U.$$

4.2.2 Preferential attachment network of Dereich

Dereich in [25] defined an alternative preferential attachment model with fitness that can be studied without a Malthusian condition. In the model a new vertex is connected to each existing vertex independently by a random number of edges, defining a multigraph.

Start with one vertex labelled one, with fitness F_1 drawn from μ and no edges. Denote the graph by $\mathcal{G}_1$. Given $\mathcal{G}_m$ with vertex set $\{1, ..., m\}$ we build $\mathcal{G}_{m+1}$ by introducing the vertex labelled $m+1$, giving it fitness F_{m+1} drawn from μ and connecting it independently to each vertex $n \in \{1, ..., m\}$ by a random number $E_{n,m+1}$ of directed edges (from vertex $m+1$ to n), which is Poisson distributed with rate

$$r_{n,m} := \beta F_n \frac{1 + \text{indegree of } n \text{ in } \mathcal{G}_m}{m},$$

where $0 < \beta < 1$ is a fixed parameter.

This model can be embedded into continuous space by letting $\tau_n = \frac{1}{\lambda}\sum_{i=1}^{n-1}\frac{1}{i}$, for $\lambda > 0$, be the time when the nth vertex is introduced and defining $Z_n(\tau_m)$, $m \geq n$ to be the indegree of vertex n prior to the establishment of vertex $m+1$, or in other words the number of edges pointing from vertices $n+1, \ldots, m$ to vertex n. In Section 4.4.2 we show that this model satisfies assumptions (A.1) to (A.4) without any Malthusian condition for $\gamma = \lambda\beta$ and ν continuous.

As an example we look at the fitness $V(t)$ of the vertex $m \in \{1, \ldots, n-1\}$ with largest degree at the time $t = \frac{1}{\lambda}\log n + C + o(1)$ when the nth vertex is introduced (where C denotes the Euler–Mascheroni constant) again in the case of Gnedenko's distribution (Example (3) in Section 1.4.1). Recall that in this case $g(x) = \frac{x}{1+x}$ and $\lambda\sigma_t = \sqrt{\lambda t + 1} - 1$. We denote by $S(t)$ the time of creation of this vertex; by Corollary 4, we have $S(t) = \sigma_t + (W + o(1))\sqrt{\sigma_t/\lambda}$ in distribution when $t \uparrow \infty$, where W is a centred

Gaussian of variance 1/2. Theorem 3 gives, in distribution when $t \uparrow \infty$,

$$\begin{aligned} V(t) &= g\Big(\lambda\sigma_t + \sqrt{\lambda\sigma_t}(W + o(1))\Big) + \mathcal{O}\Big(g'\big(\lambda\sigma_t + \sqrt{\lambda\sigma_t}(W + o(1))\big)\Big) \\ &= 1 - \frac{1}{\lambda\sigma_t} + \frac{W + o(1)}{\sqrt{\lambda\sigma_t}} = 1 - \frac{1}{\sqrt{\lambda t}} + \frac{W + o(1)}{(\lambda t)^{1/4}}, \end{aligned}$$

so that there is asymptotic normality for the fitness of the vertex of maximal degree. This is in contrast to the result in Corollary 2 (ii) for the case of μ in the maximum domain of attraction of the Weibull distribution, where $t(1 - V(t))$ converges to a Gamma distribution.

4.3 Random permutations with random cycle weights

Let $\theta \geq 0$ be a fixed parameter and suppose we are given a permutation σ of the indices $\{1, \dots, n\}$ and, for each of the k cycles of the permutation, a weight W_j, $j = 1, \dots, k$. Denote the length of the cycles by $Z_1, \dots, Z_k$. We create a permutation σ' of the indices $\{1, \dots, n+1\}$ from this as follows

- either pick one of the indices $m \in \{1, \dots, n\}$ from the jth cycle with probability $\frac{W_j}{n+\theta}$ and insert the new index into its cycle so that we have $\sigma'(m) = n+1, \sigma'(n+1) = \sigma(m)$ and $\sigma'(i) = \sigma(i)$ for all $i \neq m, n+1$;
- with the remaining probability $1 - \frac{\sum_{j=1}^{k} Z_j W_j}{n+\theta}$ the new index $n+1$ is mapped onto itself, creating a new cycle. This cycle is given a weight W_{k+1} sampled, independently of everything else, from μ.

The resulting process (σ_n) can be seen as a *disordered Chinese restaurant process*. The idea is that the cycles correspond to tables and new customers either join a table with a probability proportional to both the weight and the number of seats on the table, or sit at a new table. In the original Chinese restaurant process customers chose to sit on a table with a probability proportional to the number of seats and the probability of introducing a new table is $\frac{\theta}{n+\theta}$, see [4, p. 92]. This corresponds to all weights being equal to one in our scenario. We briefly mention that this model differs from the model of Betz, Ueltschi and Velenik on random permutations with cycle weights, as in their case the weight of a cycle is not random and instead depends on the size of the cycle, see [13].

Our analysis applies to this model for arbitrary parameter $\theta \geq 0$, we give details in Section 4.4.3. The key argument is again to find a suitable embedding into continuous-time: we find one such that

$$\tau_n = \log n + T_\theta + o(1),$$

where T_θ is a random variable depending on the parameter $\theta \geq 0$. The size $Z_n(t)$ of the nth cycle at time t is such that $(Z_n(t + \tau_n) \colon t \geq 0)$ are independent Yule processes with parameter F_n, so that the key parameters in our assumptions are $\lambda = \gamma = 1$. Our results refer to the largest cycle in the permutation and the smallest index in this cycle. To give an example of a result that follows from our analysis we look at the ratio $R(t)$ of the size of the largest and second largest cycle in the permutation at time t. If μ satisfies the assumptions (A.5), by Theorem 3, we have, for $x > 1$,

$$\lim_{t\to\infty} \mathbb{P}(R(t) \geq x) = \iiint \exp\big(- \zeta((-\infty,\infty) \times (-\infty,\infty) \times (z/x,\infty))\big)\, \zeta(\mathrm{d}s,\ \mathrm{d}f,\ \mathrm{d}z).$$

Using that $\nu(x) = \mathrm{e}^{-x}$ and $a_3 = 1$ in the first equality (as in (2.19)) below and the change of variable $v = f - \log y$ in the second, we get that

$$\begin{aligned}\zeta\big((-\infty,\infty) \times (-\infty,\infty) \times (z/x,\infty)\big)\big) &= \iint \mathrm{d}s\, \mathrm{d}f\, \mathrm{e}^{s^2 a_2 - 2f} \int_{z/x}^{\infty} \mathrm{e}^{-y\mathrm{e}^{s^2 a_2 - f}}\, \mathrm{d}y \\ &= \iint \mathrm{d}s\, \mathrm{d}v\, \mathrm{e}^{s^2 a_2 - 2v} \mathrm{e}^{-\mathrm{e}^{s^2 a_2 - v}} \int_{z/x}^{\infty} y^{-2}\, \mathrm{d}y = a_5\, \frac{x}{z},\end{aligned}$$

where a_5 is a positive constant. Hence, substituting f by $f + \log x$ in the final step,

$$\begin{aligned}\lim_{t\to\infty} \mathbb{P}(R(t) \geq x) &= \iint \mathrm{d}s\, \mathrm{d}f \int_0^{\infty} \mathrm{d}z\, \mathrm{e}^{-f} \mathrm{e}^{s^2 a_2 - f} \mathrm{e}^{-z(\mathrm{e}^{s^2 a_2 - f}) - a_5 \frac{x}{z}} \\ &= \iint \mathrm{d}s\, \mathrm{d}f \int_0^{\infty} \mathrm{d}w\, \mathrm{e}^{-f} \mathrm{e}^{-w - a_5 \frac{1}{w} \mathrm{e}^{s^2 a_2 - f + \log x}} = \frac{1}{x}.\end{aligned}$$

Similarly, if μ satisfies the assumptions (B.5), we have $\zeta\big((-\infty,\infty)\times(0,\infty)\times(z/x,\infty)\big)\big) = a_6\, \frac{x}{z}$, and hence by Theorem 1,

$$\lim_{t\to\infty} \mathbb{P}(R(t) \geq x) = \int \mathrm{d}s \int_0^{\infty} \mathrm{d}f \int_0^{\infty} \mathrm{d}z\, \alpha f^{\alpha-1} \mathrm{e}^{2s+f} \mathrm{e}^{-z\mathrm{e}^{s+f} - a_6 \frac{x}{z}} = \frac{1}{x}.$$

Note that this is in contrast to the case without disorder where the cycles have macroscopic size and the distribution of the asymptotic ratio is given by the ratio of the two largest elements in the Poisson–Dirichlet distribution.

4.4 Verification of Assumptions (A.1) to (A.4) for the different applications

In this section we prove that the examples we described above satisfy Assumptions (A.1) to (A.4).

4.4.1 Assumptions (A.1) to (A.4) for reinforced branching mechanisms

We give a general construction for the reinfoced branching process where at a birth event with probability p_{ij} we create i offspring of the same family and j new families. We denote the first and second marginal by $(p_i^{(1)})$ and $(p_j^{(2)})$ and the means by $m^{(1)}$ and

$m^{(2)}$, respectively. We can construct the model on an explicit probability space. Let

- F be a μ-distributed random variable,
- independently of F construct a continuous time jump process $Y = (Y(t) : t \geq 0)$ as follows
  - start at time 0 in state $Y(0) = 1$,
  - if Y is in state $k \in \mathbb{N}$ the next jump event follows at rate k,
  - let $0 < t_1 < t_2 < t_3 < \dots$ be the increasing sequence of times at which jump events happen,
  - at jump time t_n sample a pair $(J_n, L_n) \in \mathbb{N}_0 \times \mathbb{N}_0$ independently from (p_{ij}) and increase $(Y(t) : t \geq 0)$ by J_n (which may be zero).
- given the above let $\Pi = (\Pi(t) : t \geq 0)$ be the jump process which has a jump of height L_n (which may be zero) at time t_n.

We let $(\Omega, \mathcal{F}, \mathbb{P})$ be the countable product of the joint law of (F, Y, Π) and denote the coordinate process by (F_n, Y_n, Π_n), for $n \in \mathbb{N}$. The process $(Y_n(t)\colon t \geq 0)$ describes the creation of new family members, and the process $(\Pi_n(t)\colon t \geq 0)$ the creation of new families descending from the nth family (in a standardised time-scale). To construct our original objects on this probability space we let $\tau_1 = 0$ and $Z_1(t) = Y_1(F_1 t)$ and given $\tau_1, \dots, \tau_{n-1}$ iteratively define

$$\tau_n = \inf\{t > \tau_{n-1} : \exists m \in \{1, ..., n-1\} \text{ with } \Delta\Pi_m(F_m(t - \tau_m)) > 0\},$$

and then if $\Delta\Pi_m(F_m(t - \tau_m)) = k \geq 2$ also set $\tau_{n+k-1} = \dots = \tau_{n+1} = \tau_n$. Further let

$$Z_n(t) = \begin{cases} Y_n(F_n(t - \tau_n)), & \text{if } t \geq \tau_n, \\ 0, & \text{otherwise.} \end{cases}$$

We let $M(t) = \max\{n : \tau_n \leq t\}$ and $N(t) = \sum_{n=1}^{M(t)} Z_n(t)$. Now $(Y_n(F_n(t-\tau_n))\colon t \geq \tau_n)$ gives the sizes of the nth family, and $(\Pi_n(F_n(t - \tau_n))\colon t \geq \tau_n)$ the times of creation of the new families which descend directly from the nth family. This construction defines a reinforced branching process in a slightly more general way than in [26].

We now check that reinforced branching processes with, for some $\eta' > 0$,

$$\sum_{i,j} \mathrm{e}^{\eta'(i+j)} p_{ij} < \infty$$

satisfy Assumptions (A.1)–(A.4). The process $(M(t) : t > 0)$ is a general branching process, also known as a Crump–Mode–Jagers process, with the laws of offspring times given by the random point process $(\Pi^*(t) : t > 0)$ given by $\Pi^*(t) = \Pi(Ft)$. Assuming that there exists $\lambda > 0$, called *Malthusian parameter*, such that

$$\int_0^\infty \mathrm{e}^{-\lambda s}\mathbb{E}\Pi^*(\mathrm{d}s) = 1, \tag{4.4}$$

we can apply a strong law of large numbers by Nerman (see [59]) which shows that under an $x \log x$ condition on Π^* there exists a positive random variable W, such that

$$\lim_{t\to\infty} \mathrm{e}^{-\lambda t} M(t) = W \qquad \text{a.s.}$$

This gives us that $\log M(t) = \log W + \lambda t + o(1)$ almost surely and plugging $t = \tau_n$ yields that $\tau_n = \frac{1}{\lambda}\log n + T + \varepsilon_n$ for $T = -\frac{1}{\lambda}\log W$ and a sequence ε_n which converges to 0 almost surely.

The Malthusian condition (4.4) reads as

$$1 = \int_0^\infty \mathrm{e}^{-\lambda s}\, \mathbb{E}\Pi^*(\mathrm{d}s) = m^{(2)} \int \frac{f}{\lambda - f m^{(1)}}\, \mu(df),$$

which has a solution $\lambda > m^{(1)}$ if and only if

$$m^{(2)} \int \frac{f}{1-f}\, \mu(df) > m^{(1)}.$$

The $x \log x$ condition states that for the random varable $X = \int_0^\infty \mathrm{e}^{-\lambda s}\Pi^*(ds)$ we have $\mathbb{E}X \log^+ X < \infty$. It is easy to check that under our assumption on the moments of (p_{ij}) we even have $\mathbb{E}X^2 < \infty$ so that this condition and hence (A.1) holds.

We let $Y_n(u) = X_n(u)$ so that $\Delta_n(t) = 0$, so the convergence in Assumption (A.2) is trivially satisfied. The process $(Y_n(t) \colon t \geq 0)$ is a continuous-time Galton-Watson process with offspring distribution $(p_i^{(1)})$ and hence Assumptions (A.3) and (A.4) follow from Lemma 26 below, parts (c) and (d), respectively.

Lemma 26 (Galton–Watson process $(Y(t) \colon t \geq 0)$ with offspring distribution $(p_i^{(1)})$)**.**

(a) $\mathbb{E}[Y(t)] = \mathrm{e}^{m^{(1)}t}$.

(b) $(\mathrm{e}^{-m^{(1)}t}Y(t))_{t\geq 0}$ *is a uniformly integrable martingale.*

(c) The almost sure limit of $\lim_{t\to\infty} \mathrm{e}^{-m^{(1)}t}Y(t)$ *is an absolutely continuous random variable.*

(d) There exist $c_0, \eta > 0$ *such that* $\mathbb{P}\big(\max_{t\geq 0} \mathrm{e}^{-\gamma t}Y(t) \geq x\big) \leq c_0\, \mathrm{e}^{-\eta x}$,

Proof. (a), (b), (c) are standard and proofs can be found in Athreya and Ney [6]. Denote the martingale limit in (c) by A. For the proof of (d) note that, as in [54, Theorem 2.1], one can use the assumption on (p_{ij}) to check that there exists $\eta > 0$ such that $\mathbb{E}e^{\eta A} < \infty$. Moreover, $(\exp\{\eta e^{-\gamma t} Y(t)\} : t \geq 0)$ is a sub-martingale, by Jensen's inequality, and Doob's martingale inequality gives $\mathbb{P}(\max_{t\geq 0} e^{-\gamma t} Y(t) \geq x) \leq \mathbb{P}(\max_{t\geq 0} \exp\{\eta e^{-\gamma t} Y(t)\} \geq e^{\eta x}) \leq \mathbb{E}[\exp(\eta A)]\, e^{-\eta x}$, as required. $\square$

Our examples in Section 4.1 and Section 4.2.1 can be fitted in this framework.

- **The branching process with selection and mutation.**

  With an offspring distribution (p_k) at a birth event and mutation probability β the process becomes a reinforced branching process with offspring distribution (p_{ij}) given by
  $$p_{ij} = p_{i+j} \binom{i+j}{i} (1-\beta)^i \beta^j,$$
  so that $m^{(1)} = (1-\beta)\mathfrak{m}$ and $m^{(2)} = \beta\mathfrak{m}$.

- **The preferential attachment tree of Bianconi and Barabási.**

  This process can be embedded into continuous time as a reinforced branching process with $p_{11} = 1$ so that $m^{(1)} = m^{(2)} = 1$, see [26] for details. Here families correspond to vertices and the family size is the vertex degree. At every birth event a new vertex of degree one (equivalently a new family) is created and by establishing an edge to an existing vertex the degree of this vertex is increased by one (equivalently one existing family is getting a new member). At time τ_n the nth vertex is introduced and, for $m > n$, the degree of this vertex when the mth vertex is introduced is $Z_n(\tau_m)$.

4.4.2 Assumptions (A.1) to (A.4) for Dereich's preferential attachment network

Here we check that Assumptions (A.1) to (A.4) are satisfied for the preferential attachment network of Dereich presented in Section 4.2.2. Assumption (A.1) is straightforward for the deterministic choice

$$\tau_n = \frac{1}{\lambda} \sum_{i=1}^{n-1} \frac{1}{i} = \frac{1}{\lambda} \log n + \frac{C}{\lambda} + o(1),$$

where C is the Euler–Mascheroni constant. To show that Assumption (A.2) is satisfied we introduce a coupling of the in-degree processes $(Z_n(t)\colon t \geq 0)$ to independent Yule processes. Recall that

$$X_n(u) = Z_n\big(\tfrac{u}{F_n} + \tau_n\big).$$

Proposition 27. *There exists a coupling of the processes $(X_n(u)\colon u \geq 0)$ and a sequence $(Y_n(u)\colon u \geq 0)$ of independent Yule processes with parameter $\gamma = \beta\lambda$ such* *that,*

$$\sup_{n\in I_\kappa(t)} \mathbb{P}\big(\sup_{u\geq t} \mathrm{e}^{-\gamma u}\big|1+X_n(u)-Y_n(u)\big| \geq \varepsilon \big| (F_m)_{m\in\mathbb{N}}\big) \longrightarrow 0.$$

To prove this start with a sequence $(Y_n(u)\colon u\geq 0)$ of independent Yule processes with parameter γ. For $m\geq n+1$ we take

$$J_n(m) = Y_n(F_n(\tau_m-\tau_n)) - Y_n(F_n(\tau_{m-1}-\tau_n)).$$

We need the following lemma.

Lemma 28. *Given n there is a coupling of $J_n(m)$ and random variables $P_n(m)$, $m\geq n+1$, such that conditionally on $F_n=f$ and $\sum_{\ell=n+1}^{m-1} P_n(\ell)=k$ the random variable $P_n(m)$ is Poisson distributed with parameter $\beta f\frac{1+k}{m}$ and*

$$\sup_{n\in I_\kappa(t)} \mathbb{P}\big(J_n(m)\neq P_n(m) \text{ for some } m\geq n+1\big) \longrightarrow 0, \qquad \text{as } t\to\infty.$$

Proof. We abbreviate $Y_n^*(t)=Y_n(F_n(t-\tau_n))$ and note that $(Y_n^*(t)\colon t\geq 0)$ is a continuous time Galton–Watson process starting with one individual at time τ_n and individuals performing binary branching at rate γF_n. The coupling is now performed in two steps.

(a) For $m\geq n+1$, we let $\mathcal{E}_{n,m}$ be the event that all of the individuals alive at time τ_{m-1} have at most one descendant in the interval $[\tau_{m-1},\tau_m)$. This means that an individual existing at time τ_{m-1} can only give birth to at most one individual, which in turn does not reproduce before τ_m.

Denoting

$$\mathcal{E}_n(t) = \bigcap_{\substack{m\geq n+1\\ \tau_{m+1}<t}} \mathcal{E}_{n,m},$$

we show that

$$\sup_{n\in I_\kappa(t)} \mathbb{P}(\mathcal{E}_n^{\mathrm{c}}(t)) \to 0, \qquad \text{as } t\to\infty.$$

(b) There are random variables $J_n^*(m)$, which, conditionally on $F_n=f$ and $\sum_{l=n+1}^{m-1} J_n^*(l) = k$, are binomially distributed with parameters $1+k$ and $\beta f/m$, such that $J_n^*(m) = J_n(m)$ on $\mathcal{E}_{n,n+1}\cap\cdots\cap\mathcal{E}_{n,m}$. We can couple $J_n^*(m)$ to random variables $P_n(m)$, which given $F_n=f$ and $\sum_{l=n+1}^{m-1} P_n(l)=k$ are Poisson distributed with parameter $\beta f\frac{1+k}{m}$ such that

$$\sup_{n\in I_\kappa(t)} \mathbb{P}\big(J_n^*(m)\neq P_n(m) \text{ for some } m\geq n+1\big) \longrightarrow 0, \qquad \text{as } t\to\infty.$$

It is clear that the lemma follows from claims (a) and (b).

We now prove (a). Fix $n\in I_\kappa(t)$ and let $m>n$. Denote by $\eta=\gamma F_n$ and by W_θ an independent random variable, exponentially distributed with parameter θ. Recall that in a Yule process of rate η each particle gives birth to one offspring after an exponentially distributed waiting time with rate η, independently of everything else. Thus the probability that a fixed particle has at least one offspring in the interval $[\tau_m, \tau_{m+1})$ is equal to

$$\mathbb{P}(W_\eta \le \tau_{m+1} - \tau_m) = 1 - \mathrm{e}^{-\frac{\eta}{\lambda m}} \le \tfrac{\beta F_n}{m}.$$

Furthermore, the probability of a given particle having at least two descendants in $[\tau_m, \tau_{m+1})$ is equal to

$$\begin{aligned} \mathbb{P}(W_\eta + W_{2\eta} \le \tau_{m+1} - \tau_m) &= 1 + \mathrm{e}^{-\frac{2\eta}{\lambda m}} - 2\mathrm{e}^{-\frac{\eta}{\lambda m}} \\ &= (1 - \mathrm{e}^{-\frac{\beta F_n}{m}})^2 \le \tfrac{(\beta F_n)^2}{m^2}, \end{aligned} \tag{4.5}$$

where $W_{2\eta}$ is the minimum of two independent exponentially distributed waiting times with rate η. Using the law of total probability we can express the probability that at least one particle of $(Y_n^*(t)\colon t \ge 0)$ at time τ_m has at least two descendants in the interval $[\tau_m, \tau_{m+1})$,

$$\mathbb{P}(\mathcal{E}^{\mathrm{c}}_{n,m+1}) = \sum_{k=1}^{\infty} \mathbb{P}(\mathcal{E}^c_{n,m+1} | Y_n^*(\tau_m) = k)\mathbb{P}(Y_n^*(\tau_m) = k) \le \tfrac{(\beta F_n)^2}{m^2}\mathbb{E}[Y_n^*(\tau_m)]. \tag{4.6}$$

By Lemma 26(a), we have

$$\mathbb{E}[Y_n^*(\tau_m)] = \mathbb{E}[Y_n(F_n(\tau_m - \tau_n))] = \mathrm{e}^{\beta\lambda F_n(\tau_m - \tau_n)} \asymp (\tfrac{m}{n})^{\beta F_n}, \tag{4.7}$$

where we have used the fact that $\tau_m - \tau_n = \frac{1}{\lambda}\log(\frac{m}{n}) + O(1)$ for $m \ge n$ and n large.

We now look at n such that $n \in I_\kappa(t)$. Recall that $I_\kappa(t)$ is defined for μ in MDA(Gumbel) and in MDA(Weibull) in Equations (1.7) and (1.9) respectively. Consider n such that

$$\mathrm{e}^{\lambda(\sigma_t - \kappa\sqrt{\sigma_t})} \le n \le \mathrm{e}^{\lambda(\sigma_t + \kappa\sqrt{\sigma_t})},$$

which implies that $n \in I_\kappa(t)$ for both cases. Putting this together with Equations (4.6) and (4.7) we get

$$\begin{aligned} \mathbb{P}(\mathcal{E}^c_n(t)) &\le \sum_{m=n}^{\infty} \mathbb{P}(\mathcal{E}^c_{n,m+1}) \le \text{const.} \sum_{m=\mathrm{e}^{\lambda(\sigma_t - \kappa\sqrt{\sigma_t})}}^{\infty} \frac{(\beta F_n)^2}{m^2}\Big(\frac{m}{n}\Big)^{\beta F_n} \\ &\le \text{const.}\frac{(\beta F_n)^2}{\mathrm{e}^{\beta\lambda F_n(\sigma_t - \kappa\sqrt{\sigma_t})}} \sum_{m=\mathrm{e}^{\lambda(\sigma_t - \kappa\sqrt{\sigma_t})}}^{\infty} m^{\beta F_n - 2} \\ &\le \text{const.}\frac{(\beta F_n)^2}{\mathrm{e}^{\beta\lambda F_n(\sigma_t - \kappa\sqrt{\sigma_t})}} \int_{\mathrm{e}^{\lambda(\sigma_t - \kappa\sqrt{\sigma_t})}}^{\infty} x^{\beta F_n - 2}\,\mathrm{d}x \le \text{const.}\frac{(\beta F_n)^2}{1 - \beta F_n}\mathrm{e}^{-\lambda(\sigma_t - \kappa\sqrt{\sigma_t})}, \end{aligned}$$

which goes to zero, as $t \to \infty$. This completes the proof of (a).

To show (b), fix $n \in I_\kappa(t)$ and let $m \geq n+1$. Note that the existence of $J_n^*(m)$ binomially distributed with parameters $k+1$ and $\beta f/m$ such that $J_n^*(m) = J_n(m)$ on $\mathcal{E}_{n,n+1} \cap \cdots \cap \mathcal{E}_{n,m}$ is easy because on this event there are $k+1$ individuals alive at time τ_{m-1} and each independently produces offspring with probability $\beta f/m$.

Moreover, by Lemma 26(c), we have $Y_n^*(\tau_m) \sim \xi_n \mathrm{e}^{\gamma F_n(\tau_m - \tau_n)} = \mathcal{O}((\frac{m}{n})^{\beta F_n} \xi_n)$ almost surely, when $m \uparrow \infty$. Applying Theorem 9 of [42, Chapter 4.12] the conditional total variation distance between $J_n^*(m)$ and $P_n(m)$ satisfies

$$d_{\mathrm{TV}}(J_n^*(m), P_n(m)) \leq 2 \sum_{i=1}^{Y_n^*(\tau_{m-1})} \Big(\frac{\beta F_n}{m}\Big)^2 = \frac{2(\beta F_n)^2 Y_n^*(\tau_{m-1})}{m^2} = \mathcal{O}\Big(\frac{(\beta F_n)^2 \xi_n}{m^{2-\beta F_n} n^{\beta F_n}}\Big),$$

when $m \uparrow \infty$, where the $\mathcal{O}$-term does not depend on n. This implies that there exists a coupling of $J_n^*(m)$ and $P_n(m)$, $m \geq n+1$, such that

$$\mathbb{P}\big(J_n^*(m) \neq P_n(m) \text{ for some } m \geq n+1\big) \leq \text{const.}\, \xi_n \sum_{m=n}^{\infty} \frac{(\beta F_n)^2}{m^{2-\beta F_n} n^{\beta F_n}} \leq \frac{\text{const.}\, \xi_n}{n},$$

using again that $\beta < 1$. This implies that

$$\sup_{n \in I_\kappa(t)} \mathbb{P}\big(J_n^*(m) \neq P_n(m) \text{ for some } m \geq n+1\big) \leq \text{const.} \sup_{n \in I_\kappa(t)} \{\xi_n/n\} \leq \text{const.} \frac{\sup_{n \in I_\kappa(t)} \xi_n}{\inf(I_\kappa(t))}.$$

Note that the cardinal of $I_\kappa(t)$ is less than or equal to $2\kappa\sqrt{\sigma_t}$, and the ξ_n's are i.i.d. standard exponential random variables. Thus, by extreme value theory, we get that, in distribution when $t \uparrow \infty$,

$$\sup_{n \in I_\kappa(t)} \xi_n = \log |I_\kappa(t)| + \mathcal{O}(1) = \log(\sigma_t)/2 + \mathcal{O}(1).$$

By definitions of $I_\kappa(t)$ for μ in MDA(Weibull) and MDA(Gumbell), we also have that $\inf(I_\kappa(t)) \geq \sigma_t - \kappa\sqrt{\sigma_t}$, thus implying that

$$\mathbb{P}\big(J_n^*(m) \neq P_n(m) \text{ for some } m \geq n+1\big) \to 0 \text{ when } t \uparrow \infty,$$

which concludes the proof. □

To complete the proof of Proposition 27 we define

$$X_n^*(F_n(t - \tau_n)) = \sum_{k=n+1}^{m} P_n(k), \qquad \text{for } \tau_m \leq t < \tau_{m+1},$$

and note that the so defined processes $(X_n^*(t) \colon t \geq 0)$ have the same distribution as

$(X_n(t) \colon t \geq 0)$. Moreover,

$$\begin{aligned}\mathbb{P}\big(1 + X_n(F_n(\tau_m - \tau_n)) = Y_n(F_n(\tau_m - \tau_n)) \text{ for all } m \geq n+1\big)& \\ = 1 - \mathbb{P}\big(J_n(m) \neq P_n(m) \text{ for some } m \geq n+1\big)&\end{aligned}$$

because $X_n(F_n(\tau_m - \tau_n)) = \sum_{k=n+1}^m P_n(k)$ and $Y_n(F_n(\tau_m - \tau_n)) = 1 + \sum_{k=n+1}^m J_n(k)$. Suppose now that $\tau_m \leq t < \tau_{m+1}$ and $1 + X_n(F_n(\tau_m - \tau_n)) = Y_n(F_n(\tau_m - \tau_n))$. Then, a.a.s. when $m \uparrow \infty$,

$$\begin{aligned}|1 + X_n(F_n(t - \tau_n)) - Y_n(F_n(t - \tau_n))| &= |Y_n(F_n(\tau_m - \tau_n)) - Y_n(F_n(t - \tau_n))| \\ &= (\xi + o(1))\mathrm{e}^{\gamma F_n(t-\tau_n)} - \xi \mathrm{e}^{\gamma F_n(\tau_m - \tau_n)} \\ &\leq (\xi + o(1))\mathrm{e}^{\gamma F_n(t-\tau_n)}\big(1 - \mathrm{e}^{-\beta F_n/m}\big),\end{aligned}$$

and hence

$$\begin{aligned}&\sup_{n \in I_\kappa(t)} \mathbb{P}\big(\sup_{u \geq t} \mathrm{e}^{-\gamma u}|1 + X_n(u) - Y_n(u)| \geq \varepsilon\big) \\ &\qquad \leq \sup_{n \in I_\kappa(t)} \mathbb{P}\Big(\xi(1 - \mathrm{e}^{-\beta F_n/n}) \geq \varepsilon/2\Big) + \sup_{n \in I_\kappa(t)} \mathbb{P}\big(J_n(m) \neq P_n(m) \text{ for some } m \geq n+1\big) \\ &\qquad \longrightarrow 0 \text{ as } t \uparrow \infty.\end{aligned}$$

This completes the proof of Proposition 27 and hence of Assumption (A.2). Further, from Lemma 26(c) we see that Assumption (A.3) holds.

Finally, to prove Assumption (A.4) we fix the fitnesses (F_n) and work with conditional probabilities. Note that, by definition, the jump of $(X_n(t) \colon t \geq 0)$ at time $t = F_n(\tau_m - \tau_n)$ given $X_n(F_n(\tau_{m-1} - \tau_n)) = k$ is Poisson distributed with parameter $\beta F_n \frac{1+k}{m}$. Hence the processes $(M_m^{(n)} \colon m \geq n)$ given by

$$M_m^{(n)} := \big(1 + X_n(F_n(\tau_m - \tau_n))\big) \prod_{\ell=n+1}^{m} \big(1 + \tfrac{\beta F_n}{\ell}\big)^{-1}$$

are martingales. The scaling factor satisfies

$$f_m^{-1} := \prod_{\ell=n+1}^{m} \big(1 + \tfrac{\beta F_n}{\ell}\big) = \mathrm{e}^{\gamma(F_n(\tau_m - \tau_n))}(1 + o(1)).$$

Hence almost sure limits $M_\infty^{(n)} = \lim_{m \to \infty} M_m^{(n)}$ exist and Doob's submartingale inequality yields

$$\begin{aligned}\mathbb{P}\big(\max_{u\geq 0} X_n(u)\mathrm{e}^{-\gamma u} \geq x\big) &\leq \mathbb{P}\big(\max_{m\geq n+1} M_m^{(n)} \geq x/2\big)\\ &\leq \mathbb{E}\big[\max_{m\geq n+1} \mathrm{e}^{2\eta M_m^{(n)}}\big]\,\mathrm{e}^{-\eta x}\\ &\leq \mathbb{E}\big[\mathrm{e}^{2\eta M_\infty^{(n)}}\big]\,\mathrm{e}^{-\eta x}.\end{aligned}$$

It remains to show that there exists $\eta > 0$ such that $\mathbb{E}[\mathrm{e}^{\eta M_\infty^{(n)}}] < \infty$ or, using Fatou's lemma, that $\mathbb{E}[\exp(\eta M_m^{(n)})]$ remains bounded. Using the generating function for Poisson variables we get

$$\mathbb{E}\big[\mathrm{e}^{\eta M_m^{(n)}}\big|X_n(F_n(\tau_{m-1}-\tau_n)) = k\big] = \exp\big((1+k)(\eta f_m + \beta F_n \tfrac{1}{m}(\mathrm{e}^{\eta f_m}-1))\big).$$

Hence, using that $\mathrm{e}^{\eta f_m} - 1 \leq \eta f_m + C\eta^2 f_m^2$ for some constant $C > 0$, we get

$$\mathbb{E}\big[\mathrm{e}^{\eta M_m^{(n)}}\big] \leq \mathbb{E}\big[\mathrm{e}^{(\eta + C\eta^2 \frac{1}{m} f_m) M_{m-1}^{(n)}}\big],$$

and iterating this we get an upper bound of $\mathrm{e}^{a_{m-n}}$ for the recursion $a_0 = \eta$ and

$$a_{i+1} = a_i + Ca_i^2 \tfrac{1}{m-i} f_{m-i}, \qquad \text{for } i \geq 0.$$

As $f_l \asymp (n/l)^{\beta F_n}$ there exists $A < \infty$ such that

$$\prod_{l=n+1}^{m} (1 + C\tfrac{1}{l} f_l) \leq A \qquad \text{for all } m \geq n \text{ and } n.$$

Hence $(a_{m-n} \colon m \geq n)$ is bounded by one if $0 < \eta < 1/A$. This completes the proof of (A.4).

4.4.3 Assumptions (A.1) to (A.4) for random permutations with cycle weights

The key is again an embedding of the process in continuous time such that T_n is the time when the nth customer enters the restaurant. We let $T_1 = 0$ and define T_{n+1}, $n \in \mathbb{N}$, inductively as follows. At time T_n we start $n+1$ independent exponential clocks, one clock of parameter one for each of the n customers seated in the restaurant and one additional clock of parameter θ for the creation of additional tables. We let T_{n+1} be the time when the first of these clocks rings.

- If it is the clock corresponding to customer m sitting at table j we toss a coin with success probability W_j.
  - If there is a success the $(n+1)$st customer joins this table, resp. in the language of random permutations the element $n+1$ is inserted in this cycle

between elements m and $\sigma_n(m)$,

 – if there is no success the $(n+1)$st customer seats at a new table which, if it is the $(k+1)$st occupied table, gets weight W_{k+1}.

- If it is the clock for the creation of additional tables, the $(n+1)$st customer also sits at a new table which, if it is the $(k+1)$st occupied table, gets weight W_{k+1}.

Suppose $W_1, W_2, \dots$ are given. We note that, as required, the overall probability that a new table is created at time T_{n+1} is

$$\frac{\sum_{j=1}^k Z_j(T_n)(1-W_j)+\theta}{n+\theta} = 1 - \frac{\sum_{j=1}^k Z_j(T_n)W_j}{n+\theta},$$

where $Z_j(T_n)$ is the number of occupants at the jth table at time T_n, and the probability that the $(n+1)$st customer joins the jth table is $Z_j(T_n)W_j/(n+\theta)$. Looking at the jth table, we let τ_j be the time when it is first occupied. If at time t this table is occupied by m customers the rate at which new customers join this table is mW_j, independent of the occupancy of other tables. The processes $(Z_j(t+\tau_j) \colon t \geq 0)$ are therefore independent Yule processes with rate W_j. Hence Assumptions (A.2)–(A.4) are satisfied for $\gamma = 1$ and where $X_n(u) = Y_n(u)$, $u \geq 0$, are given by $Z_n(t) = X_n(W_n(t-\tau_n))$.

Finally, to check Assumption (A.1) we note that the process of introduction of new tables is a general branching process with immigration. The immigration process corresponds to the creation of the additional tables, which is a homogeneous Poisson process with rate θ. The point process of creation of tables by unsuccessful coin tossing is a Cox process $(\Pi(t) \colon t \geq 0)$, that is, a Poisson process with random intensity. Its intensity is given by $(1-W)Y(t)\,\mathrm{d}t$ where W has distribution μ and given W the process $(Y(t) \colon t \geq 0)$ is a Yule process with parameter W. The relevant results for general branching processes can be found in [59] with the case of branching processes with immigration treated in [64]. The crucial assumption is the existence of a Malthuisan parameter $\alpha \geq 0$ such that

$$1 = \int \mathrm{e}^{-\alpha t}\, \mathbb{E}\Pi(\mathrm{d}t) = \int \int_0^\infty (1-w)\mathrm{e}^{-\alpha t} e^{wt}\, dt\, \mu(\mathrm{d}w) = \int \frac{1-w}{\alpha - w}\, \mu(\mathrm{d}w),$$

which is always satisfied for $\alpha = 1$. As above, the $x \log x$ condition on $\int \mathrm{e}^{-t}\Pi(\mathrm{d}t)$ can be checked easily. We obtain from [59, Theorem 5.4] for general branching processes without immigration (our case $\theta = 0$) and modifications described in [64, Theorem 4.2] for the general case (stated there only for convergence in L^1) that there exists a positive random variable N_θ such that the total number $N(t)$ of customers which have arrived by time t satisfies

$$\mathrm{e}^{-t} N(t) \longrightarrow N_\theta \quad \text{almost surely},$$

from which we infer that $\tau_n = \log n - \log N_\theta + o(1)$, which is Assumption (A.1) with $\lambda = 1$.

Chapter 5

Conclusion

To conclude let us recall that our results, motivated by the Bianconi–Barabási model could be applied to different types of networks with preferential attachment with multiplicative fitness, random permutations with cycle weights, such as the disordered Chinese restaurant processes, and to branching processes with selection and mutation. We have shown that for bounded fitness distributions in two classes: MDA(Gumbel) and MDA(Weibull) we can prove results about the largest family in the competing growth processes. For both classes the size, rescaled by a function of t and γT converges to a Fréchet distributed random variable.

The CGPs framework paves way for many interesting questions, inspired by its various applications. How does the system grows when the fitness distribution is unbounded and when the Malthusian growth rate does not exist? What happens if particles are not immortal but have a random death rate, or when the mutants' fitness depend on the parent, or parents? Is there universality in the ratio of the sizes of tables in the disordered Chinese restaurant process?

Appendix

For our proofs we need the following consequences of the mean value theorem.

Lemma 29. *For all* $x \in [0,1]$*, there exists* $c_3 \in \big[g(\lambda\sigma_t), g(\lambda\sigma_t) + \frac{w\sqrt{\sigma_t}}{t} g(\lambda\sigma_t) + \frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t - \frac{1}{\gamma t}\log\big(-\frac{1}{\eta\varepsilon}\log x\big)\big]$ *such that*

$$
\begin{aligned}
&m\Big(g(\lambda\sigma_t) + \frac{w\sqrt{\sigma_t}}{t} g(\lambda\sigma_t) + \frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t - \frac{1}{\gamma t}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big)\Big) \\
&\quad = m(g(\lambda\sigma_t)) + m'(g(\lambda\sigma_t))\Big(\frac{w\sqrt{\sigma_t}}{t} g(\lambda\sigma_t) + \frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t - \frac{1}{\gamma t}\log\big(-\tfrac{1}{\eta\varepsilon}\log x\big)\Big) \\
&\quad + \frac{1}{2} m''(c_3)\Big(\frac{w\sqrt{\sigma_t}}{t} g(\lambda\sigma_t) + \frac{a_1 g(\lambda\sigma_t)}{\gamma t}\log\sigma_t - \frac{1}{\gamma t}\log\Big(-\tfrac{1}{\eta\varepsilon}\log x\Big)\Big)^2.
\end{aligned}
\tag{5.1}
$$

For $\tilde{x}_0 > 0$ *and* $\kappa > 0$ *there exist* $c_4 \in \big[g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t}), g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t}) - \kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\big]$*, such that*

$$
\begin{aligned}
&m\Big(g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t}) - \kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big) = m\Big(g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big) \\
&\quad = m\Big(g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big) + m'\Big(g(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\big)\big(-\kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big) \\
&\quad + \tfrac{1}{2} m''(c_4)\Big(-\kappa g'(\lambda\tilde{x}_0\sigma_t + \log\sqrt{\sigma_t})\Big)^2.
\end{aligned}
\tag{5.2}
$$

And finally for all $w \in [-\infty, \infty]$ *there exists* $c_5 \in [\lambda\sigma_t, \lambda\sigma_t + \lambda w\sqrt{\sigma_t} + \log\sqrt{\sigma_t}]$ *such that*

$$
g'(\lambda\sigma_t + \lambda w\sqrt{\sigma_t} + \log\sqrt{\sigma_t}) = g'(\lambda\sigma_t) + g''(c_5)(\lambda w\sqrt{\sigma_t} + \log\sqrt{\sigma_t}). \tag{5.3}
$$

Proof. These follow from the mean value theorem. □

Bibliography

[1] L. A. Adamic and B. A. Huberman. Power-law distribution of the world wide web. *Science*, 287(5461):2115, 2000.

[2] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[3] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378, 2000.

[4] D. Aldous. Exchangeability and related topics. In *Lecture notes in mathematics*, volume 1117, pages 1–198. Springer-Verlag, 1985.

[5] K. B. Athreya and S. Karlin. Embedding of urn schemes into continuous time markov branching processes and related limit theorems. *The Annals of Mathematical Statistics*, 39(6):1801–1817, 1968.

[6] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer-Verlag, New York, 1972.

[7] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42. ACM, 2012.

[8] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[9] A.-L. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1-2):173–187, 1999.

[10] A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: Statistical Mechanics and its Applications*, 281(1-4):69–77, 2000.

[11] A. Barrat and M. Weigt. On the properties of small-world network models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 13(3):547–560, 2000.

[12] A. Békéssy, P. Békéssy, and J. Komlós. Asymptotic enumeration of regular matrices. *Studia Scientiarum Mathematicarum Hungarica*, 7:343–353, 1972.

[13] V. Betz, D. Ueltschi, Y. Velenik, et al. Random permutations with cycle weights. *The Annals of Applied Probability*, 21(1):312–331, 2011.

[14] S. Bhamidi. Universal techniques to analyze preferential attachment trees: Global and local analysis. Preprint available at `https://pdfs.semanticscholar.org/e7fb/8c999ff62a5f080e4c329a7a450f41fb1528.pdf`, 2007.

[15] G. Bianconi and A.-L. Barabási. Bose–Einstein condensation in complex networks. *Physical Review Letters*, 86(24):5632–5635, 2001.

[16] G. Bianconi and A.-L. Barabási. Competition and multiscaling in evolving networks. *Europhysics Letters*, 54(4):436–442, 2001.

[17] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[18] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[19] C. Borgs, J. Chayes, C. Daskalakis, and S. Roch. First to market is not everything: an analysis of preferential attachment with fitness. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 135–144. ACM, 2007.

[20] T. Britton, M. Deijfen, and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, 124(6):1377–1397, 2006.

[21] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Computer networks*, 33(1-6):309–320, 2000.

[22] A. Broido and K. Claffy. Internet topology: connectivity of IP graphs. In *SPIE International Symposium on Convergence of IT and Communication*, pages 172–187, Denver, CO, Aug 2001.

[23] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89(25):258702, 2002.

[24] Q. Chen, H. Chang, R. Govindan, and S. Jamin. The origin of power laws in internet topologies revisited. In *Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 2, pages 608–617. IEEE, 2002.

[25] S. Dereich. Preferential attachment with fitness: Unfolding the condensate. *Electronic Journal of Probability*, 21:Paper 3, 2016.

[26] S. Dereich, C. Mailler, and P. Mörters. Non-extensive condensation in reinforced branching processes. *Annals of Applied Probability*, 27(4):2539–2568, 2017.

[27] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: Degree evolution. *Electronic Journal of Probability*, 14:1222–1267, 2009.

[28] S. Dereich and P. Mörters. Emergence of condensation in Kingman's model of selection and mutation. *Acta applicandae mathematicae*, 127(1):17–26, 2013.

[29] S. Dereich and M. Ortgiese. Robust analysis of preferential attachment models with fitness. *Combinatorics, Probability and Computing*, 23(3):386–411, 2014.

[30] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *Science*, 301(5634):827–829, 2003.

[31] S. N. Dorogovtsev and J. F. F. Mendes. Effect of the accelerating growth of communications networks on their structure. *Physical Review E*, 63(2):025101, 2001.

[32] S. N. Dorogovtsev and J. F. F. Mendes. *Handbook of Graphs and Networks: From the Genome to the Internet, chapter Accelerated growth of networks*. John Wiley & Sons, 2003.

[33] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Structure of growing networks with preferential linking. *Physical Review Letters*, 85(21):4633–4636, 2000.

[34] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin, 1997.

[35] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[36] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–61, 1960.

[37] P. Erdős and A. Rényi. On the strength of connectedness of a random graph. *Acta Mathematica Scientia Hungary*, 12:261–267, 1961.

[38] G. Ergün and G. J. Rodgers. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1-2):261–272, 2002.

[39] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 251–262. ACM, 1999.

[40] N. Freeman and J. Jordan. Extensive condensation in a model of preferential attachment with fitnesses. *arXiv preprint arXiv:1812.06946*, 2018.

[41] B. V. Gnedenko. Sur la distribution limité du terme d'une série aléatoire. *Annals of Mathematics*, 44:423–453, 1943.

[42] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes.* Oxford University Press, Oxford, second edition, 2009.

[43] J. Guare. *Six Degrees of Separation: A Play.* Vintage, New York, 1990.

[44] A. Hodgins-Davis, D. P. Rice, and J. P. Townsend. Gene expression evolves under a house-of-cards model of stabilizing selection. *Molecular biology and evolution*, 32(8):2130–2140, 2015.

[45] P. Jagers and O. Nerman. The asymptotic composition of supercritical, multi-type branching populations. *Séminaire de probabilités de Strasbourg*, 30(3):40–54, 1996.

[46] S. Janson. Functional limit theorems for multitype branching processes and generalized Polya urns. *Stochastic Processes and their Applications*, 110(2):177–245, 2004.

[47] H. Jeong, Z. Néda, and A.-L. Barabási. Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61(4):567–572, 2003.

[48] J. F. Kingman. A simple model for the balance between selection and mutation. *Journal of Applied Probability*, 15(1):1–12, 1978.

[49] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. TheWeb as a graph: measurements, models, and methods. In *International Computing and Combinatorics Conference*, pages 1–17. Springer, 1999.

[50] P. L. Krapivsky and S. Redner. Organization of growing random networks. *Physical Review E*, 63(6):066123, 2001.

[51] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical review letters*, 85(21):4629–4632, 2000.

[52] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the Web graph. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 57–65. IEEE, 2000.

[53] S. Lawrence and C. Giles. Accessibility of information on the web. *Nature*, 400(1):107–109, 1999.

[54] Q. Liu. The exact Hausdorff dimension of a branching set. *Publications mathématiques et informatique de Rennes*, (2):1–38, 1993.

[55] C. Mailler, P. Mörters, and A. Senkevich. Branching processes with reinforcement. `http://math.nsc.ru/LBRT/v1/school2016/Novosibirsk.pdf`.

[56] C. Mailler, P. Mörters, and A. Senkevich. Competing growth processes with random growth rates and random birth times. *arXiv preprint arXiv:1909.07690*, 2019.

[57] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.

[58] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.

[59] O. Nerman. On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 57(3):365–395, 1981.

[60] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(2):025102, 2001.

[61] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[62] M. E. J. Newman. Random graphs as models of networks. In *Handbook of Graphs and Networks*, pages 35–68. John Wiley & Sons, 2005.

[63] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001.

[64] P. Olofsson. General branching processes with immigration. *Journal of Applied Probability*, 33(4):940–948, 1996.

[65] R. Pemantle et al. A survey of random processes with reinforcement. *Probability Surveys*, 4:1–79, 2007.

[66] D. J. d. S. Price. Networks of scientific papers. *Science*, 149(3683):510–515, 1965.

[67] D. J. d. S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5):292–306, 1976.

[68] S. I. Resnick. *Extreme values, regular variation and point processes.* Springer, 1987.

[69] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[70] R. Solomonoff and A. Rapoport. Connectivity of random nets. *The bulletin of mathematical biophysics*, 13(2):107–117, 1951.

[71] R. Van Der Hofstad. *Random graphs and complex networks*, volume 1. Cambridge university press, 2016.

[72] D. J. Watts. The dynamics of networks between order and randomness. *Small Worlds*, 1999.

[73] D. J. Watts. Networks, dynamics, and the small-world phenomenon. *American Journal of sociology*, 105(2):493–527, 1999.

[74] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.